2017

Travel Demand Modeling for London, Ontario, Canada

Rahaf Husein
University of Windsor

TRAVEL DEMAND MODELING FOR LONDON, ONTARIO, CANADA

By

Rahaf Husein

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Science
at the University of Windsor

Windsor, Ontario, Canada

2017

TRAVEL DEMAND MODELING FOR LONDON, ONTARIO, CANADA

by

Rahaf Husien

APPROVED BY:

W. Anderson
Department of Political Science

C. Lee
Department of Civil and Environmental Engineering

J. Tofflemire, Special Member
Department of Civil and Environmental Engineering

H. Maoh, Advisor
Department of Civil and Environmental Engineering

May 5th, 2017

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Travel demand modeling is one of the key areas in transportation planning and engineering. Traditionally, it has been based on four inter-connected modules: trip generation, trip distribution mode choice, and traffic assignment. While the traditional approach remains popular among practitioners, it has been criticized widely due to its zonal aggregate nature. There has been a shift towards using micro-based models that use households and members of households as the units of analysis in lieu of the traffic analysis zones. This thesis contributes to advancing this micro-based paradigm by studying travel demand in the London Census Metropolitan Area (CMA), Ontario. It does so by developing an improved four-step travel demand model using a recent household travel survey that was collected in the year 2009. The focus will be as follows: first, compare various techniques that could be used to model trip generation (i.e., regression, cross-classification, discrete choice, and count models) at the micro-level. Also, compare the predictive ability of these micro-models against conventional zone-based models. Second, apply advanced geo-spatial methods and statistical techniques to model trip distribution using micro-data from the London Household Travel Survey (LHTS). To date, trip distribution in the four-stage model has relied on the gravity approach, which is too simplistic to capture the real complexities of spatial interactions between the traffic analysis zones forming an urban area. Also, its aggregate nature does not allow it to adequately capture the interaction of the traveler's socio-economic characteristics with the attributes of alternative destinations. The results will allow us to devise an improved four-stage model that makes use of a conventional Household Travel Survey. Here, advanced techniques will be employed to improve the predictability of these models.

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Dr. Hanna Maoh for his constant guidance and encouragement. His passion in transportation and his hard-working attitude inspired me to do my best. I am also extremely thankful for the extra time Dr. Maoh spent to help me prepare and submit conference proceeding, presentations, and scholarship applications. I would also like to express my gratitude to the City of London for the London data and Dr. Tor Oiamo for helping us get access to the survey and the data used in the analysis. In addition, I would like to thank my committee members Dr. Chris Lee, Dr. William Anderson and Mr. John Tofflemire for taking the time to review my work and provide insightful comments which ultimately strengthened my thesis, and the chair of the committee, Dr. Rajesh Seth.

I especially thank Mr. Shakil Khan for his expertise in managing the Transportation System Innovation lab and for always helping me when I had troubles with my workstation. Additionally, I would like to thank my fellow lab mates, it was a pleasure working with and learning from such a great group of highly talented people. I am extremely grateful for Aya Hagag, Amal Ghamrawi, Terence Dimatulac, Kevin Gingerich, and Georgina Madar for their continuous help and support throughout my work. I am also thankful for my best friends Sara Alhasan, Aya Hagag and Amal Ghamrawi for listening to my complaints and for their constant encouragement. As for the financial support, I am grateful for the Natural Science and Engineering Research Council of Canada (NSERC), Ontario Graduate Scholarship (OGS), and RC Spencer Inc. Scholarship for supporting this research through a CGS Master's Scholarship.

Lastly, I must express my very profound gratitude to my parents and siblings for providing me with unfailing support and continuous encouragement. Without their love and support, this accomplishment would not have been possible. Thank you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Overview

Over the past six decades, intense highway development encouraged suburbanization in many Canadian cities. Consequently, over time, the form of cities evolved from compact to sprawled, thus giving rise to increased commuting times. Coupled with the continuous increase in population size, demand for personal travel and longer commutes will continue in the future. The sprawled urban planning process in Canada makes accessing jobs and retail locations by non-motorized modes inconvenient and time consuming in many cases. Therefore, driving is found to be the most favored mode of transportation with a mode share of about 74% (The Vanier Institute of the Family, 2013). According to Environment Canada (2016), Canada's total greenhouse gas (GHG) emissions in 2014 was 732 Megatonnes (Mt) of carbon dioxide equivalent (CO_2 eq), in which transportation accounts for 171 Mt CO_2 eq (about 23% of total emissions). With current policies and reliance on motorized vehicles, GHG emissions are expected to rise further. Also, with our current driving habits and with driving being the dominant mode choice, planning for future growth is essential.

Transportation planning processes have been used intensively to estimate future travel demand in the urban context. As a result, transportation models are used to capture certain aspects of the urban planning process. Researchers and policy makers can then use these models to make informed decisions on the future development and management of transportation systems. Since transportation has significant effects not only on the environment and mobility, but also on land use, economic development, government finance, and the quality of life in general, creating high quality transportation services becomes vital.

The majority of the efforts in the literature have been focused on developing aggregate zone based models (TMIP, 2014). In this framework, the urban area is divided into a finite number of zones known as traffic analysis zones (TAZs) that form the units of analysis in the model. However, with the need to capture the travelers' behavior, there has been a shift to developing micro-based models that make use of households and individuals as the unit of analysis. The zone based approach has been criticized due to the lack of the behavioral realism needed to capture the actual travel demand observed in the urban area. In such models, behavior is estimated for the zones, however, activities are made by the individuals and households, not the zones. Another major drawback is that when aggregate models are estimated, it is assumed that land use is homogenous in the TAZs. The previous is not always feasible as predetermined zoning systems do not consider the ongoing spatio-temporal changes of land use. In addition, when considering areas such as Central Business Districts (CBD) and mixed-use developments, land use is not homogenous and is very difficult to be isolated within a TAZ. Therefore, the previous assumption is not valid and using a zonal framework in such cases will produce inferior estimates to project future travel demand. This is particularly the case in the trip generation (i.e. production and attraction) and trip distribution models (i.e. spatial interactions).

The history of travel demand modeling has been dominated by the traditional four-step model (FSM). Traditionally, it has been based on four inter-connected modules: trip generation, trip distribution mode choice, and traffic assignment (for more information; Zhong et al., 2015). The main reasons for such practice are the simplicity of applying the model and the relatively inexpensive and low effort in data collection to implement the model. However, as travel is always derived from the need to engage in activities, it can be modeled either with trip-based or activity-based methods. The conventional trip-based model divides the tour made by an individual to different trips belonging to different purposes (i.e., work, school, shopping, and social and recreational) and each trip purpose is modeled separately (Hanson, 1980; Horner and O'Kelly, 2007; Krizek, 2003). This method requires data derived from Household Travel surveys such as household and socio-economic data on the individuals living in the household (i.e., income, household size, workers and employment type, and vehicle ownership).

Activity-based models (see for example: Axhausen and Graling, 1992; Castiglione et al., 2015; Ettema and Timmermans, 1997; Griesenbeck and Garry, 2007; Kitamura et al., 2000) trace the activities of a traveler in a chain. In other words, these models can incorporate and combine all the decisions the traveler makes. Consequently, they are found to be very realistic and close to reality. Although recent studies have been moving towards the activity-based methods, it should be moted that these models use travel diaries, which are not easy to collect, too costly, and difficult to come across in general. Travel diaries typically include information on the type of activity performed, location, start time, duration, mode of travel, departure time, and arrival time in addition to household and socioeconomic information.

The need for travel diaries to develop activity-based models makes these models data hungry and computationally cumbersome. As a result, the availability of the data and the objective of the analysis can dictate the type of techniques to be used to study a specific aspect of urban travel demand. Given the availability of conventional household surveys in most cases, advanced techniques could be employed to model trips generated by household in an attempt to improve the predictability of the four-stage travel demand models.

1.2 Statement of the Problem

This research strives to advance the micro-based paradigm by studying travel demand in the London Census Metropolitan Area (CMA). It does so by using a recent household travel survey that was conducted in the year 2009. The focus will be as follows: first, compare various techniques that could be used to model trip generation (i.e., regression, cross-classification, discrete choice, and count models) at the micro-level. Also, compare the predictive ability of these micro-models against conventional zone-based models. Second, apply advanced geo-spatial methods and statistical techniques to model trip distribution using micro-data from the London Household Travel Survey (LHTS). To date, trip distribution in the four-stage model has mostly relied on the gravity approach, which is too simplistic for capturing the complexities of spatial interaction within a travel demand model. The results will allow us to devise an improved four-stage model that makes use of a conventional Household Travel Survey. Here, advanced techniques will be employed to improve the predictability of these models.

1.3 Objectives

The primary objectives of this research project are:

1) developing an improved four-step travel demand model,
2) compare various techniques that could be used to model trip generation at the micro-level. Also, compare the predictive ability of these micro-models against conventional zone-based models, and
3) apply advanced geo-spatial methods and statistical techniques to model trip distribution of individual travelers.

1.4 Thesis Outline

The remainder of this thesis is organized as follows. The next chapter provides an overview of previous studies on trip generation and trip distribution. The third chapter highlights the study area and the datasets used in the analysis. This is followed by a fourth chapter to discuss the methods of analysis used in this study. The fifth chapter highlights and discusses the empirical results, while the last chapter provides conclusions and recommendations.

CHAPTER 2: LITERATURE REVIEW

This chapter reviews the literature on travel demand modeling in the urban context. Emphasis is placed on households who, as will be highlighted, are found to express their transportation needs and preferences through various decisions.

2.1 Trip Generation

2.1.1 Factors Influencing Trip Generation

The current literature makes a persuasive case on the factors influencing trip generation for an average weekday. The reviewed studies highlight the relationship between trip frequency and the following measures: household size, employment, gender, age, household income, and mobility tools. In addition, some studies emphasize the importance of separating the generated trips based on the purpose of the trip. For instance, previous studies show that increase in household size has a positive impact on the total trip frequency for both all-purpose (Badoes & Chen, 2004) and non-work trips (Huntsinger et al., 2013; Jang, 2005). The previous is expected as the number of activities household members engage in increase with the size of the household. Similarly, employment has also shown to positively impact the total number of generated all-purpose trips (Badoe, 2007; Badoe & Chen, 2004; Roorda et al., 2010), work trips (Chang et al., 2014; Huntsinger et al., 2013; Páez et al., 2006) and non-work trips (Jang, 2005). In conclusion, the previous seven studies prove that increase in economically active members in a household not only increases work trips, but also non-work trips, as the members of these households tend to be more active and engaging in social life. Gender also appears to impact trip generation. Past studies have shown that males have a positive influence on total trips (Badoe, 2007; Badoe & Chen, 2004) while females have a positive impact on shopping and social trips (Fox & Patruni, 2015). This is expected as in a multi-person household, females are typically more likely to engage in the traditional roles of shopping and social activities. Furthermore, a study conducted by Nobis et al. (2004), presented at the Transportation Research Board (TRB), highlights that the gender difference in travel patterns is linked to employment status, household structure, child care, and maintenance tasks. From their findings, one can conclude that for single families, the travel patterns of men and women are very similar. However, for multi-person households without children, differences in the travel behavior of men and women are noticed. These differences become very evident when considering multi-person households with children.

The current evidence on the relationship between age and trip generation is inconsistent (Badoe, 2007; Huntsinger et al., 2015; Jang, 2005; Páez et al., 2006; Roorda et al., 2010). Nevertheless, most studies show that the relationship between age and trip generation is non-linear. For instance, an examination by Badoe (2007) of age on total trip frequency shows that young people (11-17 years) are more likely to engage in more trips, and the number of generated trips decreases as age increases, with the lowest trips generated for seniors. Another study by Roorda et al. (2010) reports that the total trips generated per household in Hamilton, Toronto and Montreal is highest for the young age group (<20 years) and the working age group (36-50 years). Finally, Páez et al. (2006) highlights that for work trips, trip rate is relatively low for the young age group (<20 years), and increases with age reaching a peak for the working age group (34-50 years). The impact of age then decreases for the pre-retirement and senior age groups. As for the non-work trips, the number of generated trips decreases rapidly for individuals in the pre-retirement and senior age groups (Páez et al., 2006).

Finally, household income and mobility tools have a positive impact on the trip frequency of work and non-work trips (Badoe, 2007; Badoe & Chen, 2004; Chang et al., 2014; Jang, 2005; Páez et al., 2006; Roorda et al., 2010). Household income reflects the degree of economic activity by a household. Also, the availability of mobility tools (e.g. vehicle ownership, driver license and transit accessibility) reflects the degree of accessibility of a household. Therefore, increase in household income and accessibility (i.e., mobility tools) suggests that a household is actively participating in economic activities (i.e., increase in work trips). Consequently, a household then becomes more active in participating in shopping and social/recreational trips.

2.1.2 Modeling Techniques

The techniques used to model trip generation in past studies can be categorized in four groups; regression, category analysis, discrete choice, and count models. The traditional models that have been widely employed in empirical studies are linear regression models (Badoe, 2007; Chang et al., 2014) and category analysis (Chang et al., 2014). Although they have shown acceptable performance from a planning perspective, there are limitations to these methods. For example, regression models have three main limitations; 1) the likelihood of negative trip rates, 2) the continuous dependent variable, 3) and the lack in incorporation of traveler behavior theory (Chang et al., 2014). First, these models assume a normal distribution for the disturbance of trip rates; therefore the dependent variable may be estimated as a negative value. Second, the dependent variable (the number of trips) is treated as a continuous random variable even though it is discrete in nature. Third, regression models simply match a statistical relationship between the independent and dependent variables. As for category analysis, although it is seen to have an advantage over regression models (Chang et al., 2014), this technique has its own limitations. They require large sample sizes to reduce uncertainty for the cell-by-cell calculation, hence incurring high cost and time.

The later limitations can be overcome using discrete choice and/or count data models; ordered logit/probit (Badoe, 2007; Chang et al., 2014; Huntsinger et al., 2015; Roorda et al., 2010; Páez et al., 2006), Poisson (Badoe, 2007; Chang et al., 2014; Jang, 2005), negative binomial (Badoe, 2007; Jang, 2005), and zero inflated models (Jang, 2005). Ordered logit and probit models are regression models that consider the ordinal nature of the dependent variable. The difference between the logit and probit models lies in the distribution of the random variables (i.e., error terms). For logit models, the errors are assumed to follow the Gumbel distribution, whereas in probit models, the errors are assumed to follow the normal distribution. On the other hand, count models like the Poisson regression model is often used for modeling count data when the observations do not suffer from over-dispersion. To overcome over-dispersion in the count data, the negative binomial models can be used. As for the zero inflated models, they are typically considered to correct for excess zero values in the dependent variable. Therefore, the model estimates two equations; one for the count model and one choice model to predict the share of zeros vs non-zero values in the dependent variable.

Badoe (2007) and Chang et al. (2014) show that the performance of the different modeling techniques differs from one dataset to the other. Some studies found that traditional modeling techniques do a better job predicting in the base year. That is, predictions with these models were associated with less error compared to other complex techniques (Badoe, 2007; Chang et al., 2014). On the other hand, one should keep in mind that although some approaches may lead to better replication of observed travel patterns, they might not necessarily lead to better forecasts. Accordingly, the predictive ability of various models must be validated by a comparison of the observed and predicted trip rates. Some validation techniques that have been used include; correlation index, root-mean-square error, and coincidence ratio, as will be explained in the methods of analysis chapter.

While statistical modeling appears to be the mainstream approach, other less commonly used methods that could be found in the literature include the neural networks approach. Arliansyah and Hartono (2015) estimated a trip attraction model using the radial basis function neural networks and compared it to a linear regression model. The study considered seven explanatory variables; population size, number of schools, number of students, number of teachers, areas of school buildings, number of offices, and number of houses. After estimating a regression and a radial function basis models for comparison purposes, the results were in favor of the radial function basis model as it provided better predictions. The study also concluded that the number of students, number of teachers, total areas of school buildings, and number of offices are important attributes in estimating trip attraction.

2.2 Trip Distribution

When considering the four-step model, the mode choice and network assignment components have been well addressed in past research efforts compared to trip distribution. As it is, trip distribution for the most part still makes use of the traditional gravity model due to its simple structure and limited data requirements. However, the gravity model is too simplistic to capture the real complexities of spatial interactions between the traffic analysis zones forming an urban area. Also, its aggregate nature does not allow it to adequately capture the interaction of the traveler's socio-economic characteristics with the attributes of alternative destinations. To overcome these limitations, a number of recent studies considered the multinomial or nested logit frameworks to model destination choice behavior (Shobeirinejad et al., 2013; Scott and He, 2012; Timmermans, 1996; Wang, 2011).

Most of the current studies on destination choice assume that the decision maker, either a household or an individual, is faced with a universal choice set of known destinations. In other words, they assume that a decision maker is knowledgeable about all the destinations in the study area. Since such an assumption is most likely incorrect, bias is introduced in the parameter estimation of the utility function resulting in inaccurate predictions. As a result, a more realistic choice sets need to be introduced in such models to eliminate any potential bias. The universal choice set can be constrained using different measures/metrics when the true choice set is unknown to the analyst. For instance, the choice set could be constrained using accessibility measures (i.e., travel time and travel distance), socio-economic status, or the real travel activity space. It is also important to note that, in such models, the destination alternatives are typically aggregated to the level of the traffic analysis zones (TAZs).

In an earlier study by Pozsgay and Bhat (2000), the feasible choice set for each individual was created by simply adding nine non-chosen alternatives randomly selected from the universal choice set and adding them to the chosen alternative. Each choice set was then made up of ten destinations in which one is the individual's actual destination. Using these choice sets Multinomial logit (MNL) trip distribution models were estimated at the disaggregate level of the decision-maker. The study tested the relationship between level of service measures and zonal and socio-demographic attributes on trip attraction. While the authors made use of the discrete choice modeling framework, their modified choice set is not a constrained one. In fact, the random sampling of alternatives is expected to produce similar estimates to those produced if the full choice set was used (McFadden, 1978). The only advantage of the random sampling is the reduction in computational needs when the universal choice set is very large.

A recent study by Kim and Lee (2017) considered the reduction of the destination choice set using the random sampling technique employed by Pozsgay and Bhat (2000) and compared it to a stratified importance sampling approach. The first approach assumes that each alternative has an equal probability of being chosen. The second approach, on the other hand, assumes that each alternative's characteristics determine its selection probability. Kim and Lee (2017) employed the two sampling approaches to draw the non-chosen alternatives to be included in each choice set. The shopping alternatives introduced in their study were also aggregated into the level of traffic analysis zones based on spatial similarities and feasibility analysis. In addition, for the stratified importance sampling approach, *Moran's I* was employed as a measure of spatial correlation as it detects spatial clustering patterns in geographic analysis. They concluded that models based on the stratified importance sampling approach were more accurate and matched the actual results better than the random sampling approach.

Similarly, the study by Scott and He (2012) highlighted the need for a realistic representation of available shopping opportunities at the TAZ level when modeling for the 818 shopping destinations in their sample. They used the potential path area (PPA) technique that determines an individual's destination choice set given their activity schedule and the spatial distribution of specific store types. For each shopping trip, a GIS-based algorithm was used to generate the TAZ-based PPA. Therefore, an individual's constrained choice set was created using the actual chosen zone and nine randomly selected non-chosen zones within the TAZ-based PPA. As for an individual's unconstrained choice set, it was created by the actual chosen zone and nine randomly selected non-chosen zones from the 817 TAZs. Using the constrained and unconstrained destination choice sets, Scott and He estimated Multinomial logit trip distribution models at the disaggregate level of the decision-maker. They concluded that the unconstrained model overestimated the effects of the explanatory variables and was associated with larger standard errors.

The majority of papers included in this review gathered primary data through the use of household travel surveys and/or travel diaries, as illustrated in Table 2-1. Different explanatory variables are considered in trip distribution analysis. These include variables on the destination's characteristics (i.e., size, parking, level of service, etc.), accessibility measures (i.e., travel time, public transport access, and the traveler's socio-demographic characteristics (i.e., age, gender, income, and employment).

Table 2-1 Review of destination choice models in previous literature

Author(s)	Study Area & Data	Number of Observations	Size of Destination Choice Set	Modeling Technique	Explanatory Variables
Timmermans (1996)	• Eindhoven, Netherlands • Stated Preference Survey	167 Shopping Trips	--	MNL	Shopping Centre characteristics (size, price, and, parking), accessibility (travel time and distance), and mode choice.
Pozsgay and Bhat (2000)	• Dallas, USA • 1996 Dallas-Fort Worth Metropolitan Area Household Activity Survey	7,770	10 (random sampling approach)	MNL	Zonal attributes and socio-demographic characteristics (age, number of cars, income, and employment).
Wang (2011)	• Toronto, Canada • 2001 Transportation Tomorrow Survey • 2002-2003 Computerized Household Activity Scheduling Elicitor Survey	--	(Time-space prism)	Nested Logit (estimated sequentially)	Income, log (number of retail stores), log (retail floor space), Travel Time for Auto x Gender.
Scott and He (2012)	• Louisville, USA • 2000 Travel Diary Survey	616 Shopping Trips	10 (Potential Path Area)	MNL	Activity duration, socio-demographic characteristics (income, employment, age-categories), and store type.
Shobeirinejad et al. (2013)	• Queensland, Australia • 2009 South East Queensland Household Travel Survey	--	--	MNL	Attractiveness of destination (size, level of service, etc.), accessibility (travel cost, parking, public transport, walking and cycling access), travel characteristics, and the nature of shopping trips.

CHAPTER 3: STUDY AREA AND DATA DESCRIPTION

3.1 Study Area

The analysis in this research is focused on the London CMA located in Southwestern Ontario, Canada as illustrated in Figure 3-1. London occupies approximately 2,665.62 square kilometers of Canada's land (Statistics Canada, 2015). According to the most recent Canadian census, the population of the London CMA in 2011 was 474,786 living in 195,055 dwellings and 214,545 jobs in the year 2011.

3.2 Data Description

The data that will be used for this analysis was acquired from two main sources: 2011 Canadian census and the London Household Travel Survey. The Canadian census data provide demographic information such as the population and employment numbers, and information on family structure and household count by dwelling type for each census tract. The London Household Travel Survey provides information on the traveler's socioeconomic characteristics such as age, gender, employment status, dwelling type, vehicle ownership, and transit accessibility. As for the trip-related information, it includes the locations of the trip's origin and destination.

In addition, the travelers were asked to reveal the mode used to make the trip and the purpose of the trip. Here, nine travel modes; auto drive, auto passenger, London transit, chartered bus, school bus, taxi, motorcycle, bicycle, and walk are recorded. The survey also classifies the trip based on nine purposes; work, work related, school, pick up/drop off passenger, shopping, social and recreational, personal business, returning home, and other. For modeling purposes, these categories were collapsed into five main groups; work, school, shopping, social and

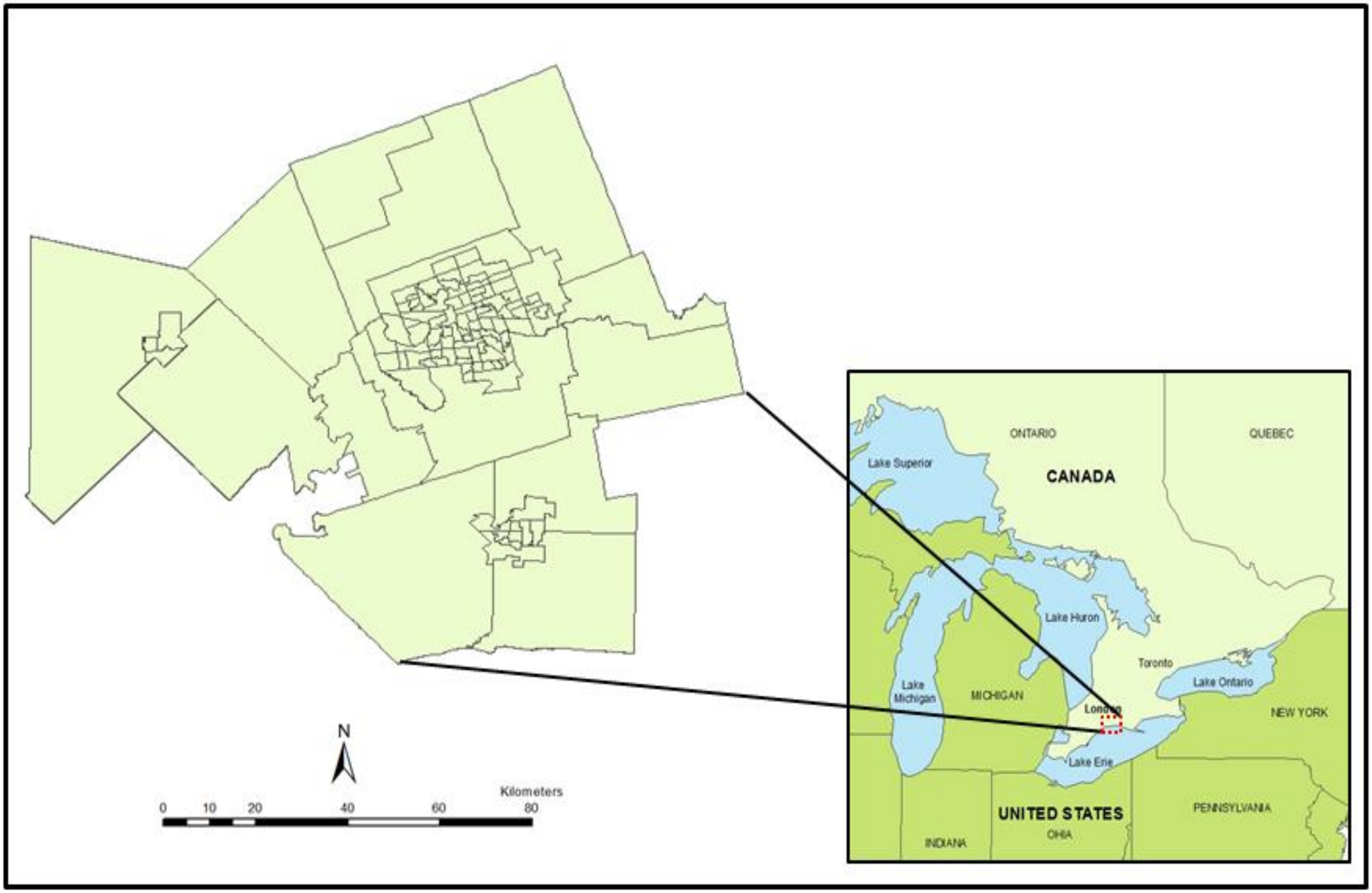

Figure 3- 1 Study area in the regional context

recreational, and other. The “other” trip purpose category includes trips that are classified in the following subcategories: personal business, pick up/drop off passenger and other. The survey is expanded using data from the Canadian census and population synthesis techniques, as will be highlighted in the next sub-section.

Out of all the work and non-work trips in the dataset, over 51% were home-based work trips, while school trips accounted for 19%. On the other hand, shopping and other trip purposed accounted for about 15% and 10% respectively. As for social and recreational trips, they pertained to about 5%. Therefore, the non-work trips pertained to about 49% of the total trips for the London CMA. The trip distribution by purpose is presented in Figure 3-2.

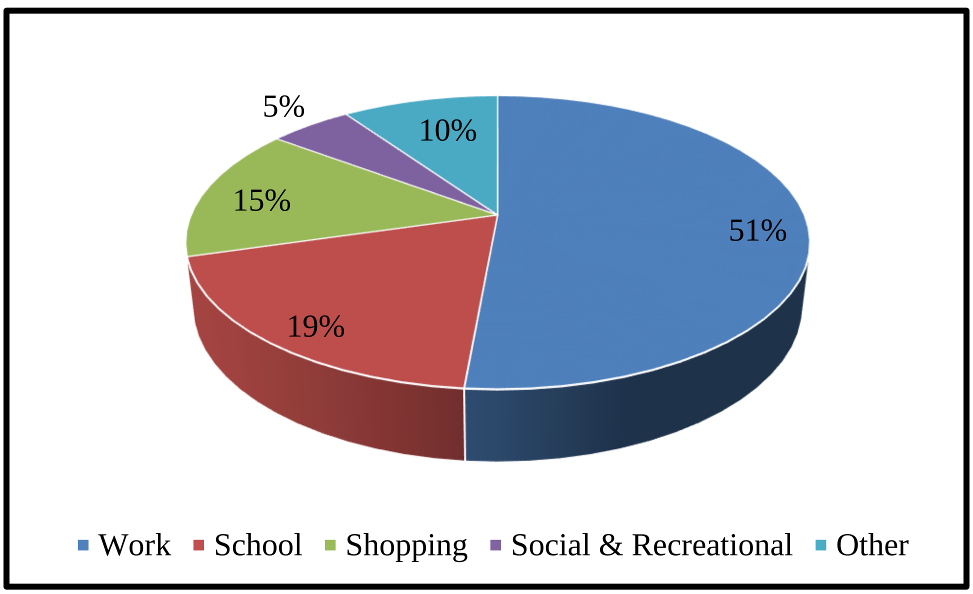

Figure 3-2 Trip distribution by trip purpose

As for the travelers’ socio-economic characteristics, females account for about 51% of the sample. Travelers under the age of 20, and aged 35-49 each represent 24% of the total sample followed by travelers aged 20-34 with 20%, then those aged 50-64 with 17% and 65+ with about

15% of the total sample. More than twenty-seven percent of the surveyed travelers live in apartments. Households with three occupants constitute nearly 35% of the sample, while 28% of the dataset consists of households with two occupants.

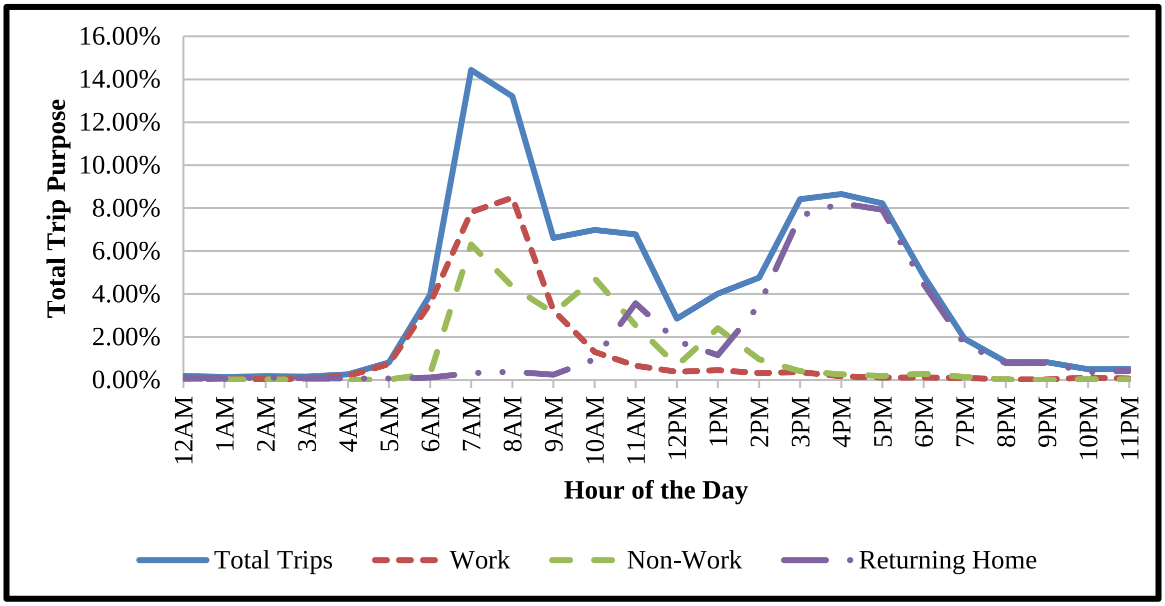

Figure 3-3 Trip distribution by trip purpose for the different hours of the day (1)

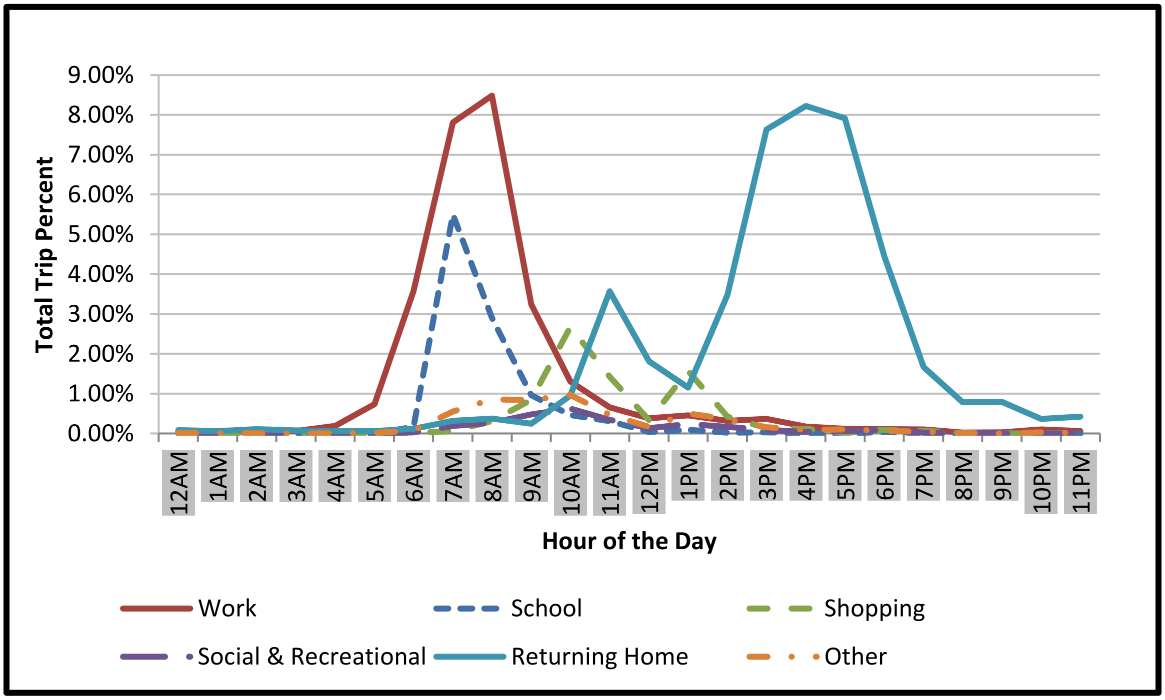

Figure 3-4 Trip distribution by trip purpose for the different hours of the day (2)

Figures 3-3 and 3-4 show the temporal distributions of the trips by trip purpose for an average weekday. Like most metropolitan areas (Saw et al., 2015), the London CMA strains the transportation network during the peak periods. Work and school trips are found to take place in the AM peak period (6-9 am) as expected. Shopping trips, on the other hand, take place in the AM-off peak and in the MID day period. Whereas social and recreational trips mostly take place in the AM-off peak period. As for the returning home trips, for large part, they take place in the PM peak period, but some take place in the MID day peak, most likely part time employees.

CHAPTER 4: METHODS OF ANALYSIS

The methods followed in this research are summarized in this chapter. Synthetic population techniques are first employed to micro-simulate the list of all individual households living in a census tract in the London CMA. This process consists of two consecutive steps that are explained in the second sub-section. Using the synthetic population list and the London Household Travel Survey (LHTS), micro-analytical trip generation and distribution models will be estimated.

4.1 Expanding the LHTS

The Combinatorial Optimization (CO) technique is used to synthesize a disaggregate list of households with attributes, which when aggregated conform to predefined zonal totals provided by the 2011 Canadian census. To reach the optimal solution, the simulated annealing (SA) approach was employed to execute the CO problem. For more information on the simulated annealing approach in the context of CO method see Williamson et al. (1998). This process was repeated multiple times to confirm the consistency of the synthesized populations.

To generate a representative aggregate cross-tabulation, as an input to the synthesizing process, zonal totals from the Canadian census were used. The cross-tabulations included information on gender, age category, and dwelling types per census tract. Next, a micro-sample of more than 6,200 households were extracted from the survey responses were information on gender, age, and dwelling type are stated. The CO method was then used to create a list of more than 195,000 households, where each household in the sample has information on the members of the household (gender and age category) and dwelling type. Each synthesized household was linked directly to a household in the micro-sample, and as such the information needed on vehicle ownership, employment, mode choice, etc. were assigned. The obtained results of the synthesized population is validated against the Census data and the results are summarized in Table 4-1.

For modeling purposes, a 5% sample (about 10,000 households) was randomly selected from the synthesized population of 195,000 households using the stratified sampling technique. The population was divided into groups based on household size (1, 2, 3, 4, 5, and 6 or more members), and a random sample was obtained from each group based on their percentage in the entire population. This was done to ensure that the sample used is representative of all groups in the entire population.

4.2 Modeling Trip Generation

4.2.1 Model Formulation

Table 4-2 lists the variables used in the analysis. These variables were inspired by the information found in the literature on the factors affecting work and non-work trip generation. Starting with socio-demographic variables, the young and senior age groups are expected to generate the least number of work trips compared to other population age groups. The highly economically active age group (35-49 years of age) is expected to generate the most work trips. As for non-work trips, it is expected that as age increases, the non-work trip rate will also increase but in a non-linear fashion. Therefore, five variables representing different age groups in a household are used and *Age1* is chosen as the reference category (RC). Besides age, gender is also observed to affect trip generation. Males are observed to generate more work trips, whereas females are observed to generate more non-work trips as they are more socially active compared to males.

Table 4-1 London population synthesis validation results

Categories		Correlations	Categories		Correlations
Age	<15 yrs	0.96	**Age**	<15 yrs	0.96
Male x Age Categories	Male (15-19)	0.94	**Female x Age Categories**	Female (15-19)	0.95
	Male (20-24)	0.78		Female (20-24)	0.83
	Male (25-29)	0.89		Female (25-29)	0.90
	Male (30-34)	0.86		Female (30-34)	0.86
	Male (35-39)	0.89		Female (35-39)	0.90
	Male (40-44)	0.92		Female (40-44)	0.94
	Male (45-49)	0.91		Female (45-49)	0.95
	Male (50-54)	0.88		Female (50-54)	0.92
	Male (55-59)	0.87		Female (55-59)	0.91
	Male (60-64)	0.87		Female (60-64)	0.86
	Male (65-69)	0.82		Female (65-69)	0.81
	Male (70-74)	0.82		Female (70-74)	0.81
	Male (75-79)	0.76		Female (75-79)	0.67
	Male (80-84)	0.70		Female (80-84)	**0.65**
	Male (85+)	**0.43**		Female (85+)	**0.59**
	Total Male	0.99		Total Female	0.99
Mode Choice (Work Trips)	Auto – Driver	0.94	**Household Size**	Hhld_1	0.99
	Auto – Passenger	0.77		Hhld_2	0.99
	Public transit	0.67		Hhld_3	0.99
	Walked	**0.41**		Hhld_4	1.00
	Bicycle	**0.13**		Hhld_5	1.00
	Other	**0.44**		Hhld_6+	1.00
Employment	Employed	0.97	**Dwelling Type**	Apartment	0.99
	Not Employed	0.76		House	0.99

In addition, by exploring the temporal distributions of the trips by trip purpose for an average weekday for London CMA, we concluded that like most metropolitan areas, the transportation network in the London CMA strains during the peak periods. Work trips are found to take place in the AM peak period (6-9 am). Whereas for non-work trips, they mostly take place in the AM-off peak and in the mid-day period. As a result, dummy variables representing the hour of the day the trip took place in are also considered. Finally, a dummy variable representing the social trips are also used to distinguish the shopping from social trips in the non-work trips analysis.

Table 4-2 List of variables used in trip generation analysis

Variable	Description	Expectation
Age 1	The population of less than 20 years in a household	RC
Age 2	The population of 20 to 34 years in a household	+
Age 3	The population of 35 to 49 years in a household	+
Age 4	The population of 50 to 64 years in a household	+
Age 5	The population of 65 years and older in a household	-/+*
Males	The population of males of 15 years and older in a household	+
Vehicles	The number of vehicles owned in a household	+
Social	1 if purpose of the non-work trip is social and recreational; 0 otherwise	+
Social × Females	An interaction term between non-work trip purpose (social and recreational) and the population of females of 15 years and older in a household	+
Dummy1	1 if the work trip took place between 6 and 7 am; 0 otherwise	+
Dummy2	1 if the work trip took place between 7 and 8 am; 0 otherwise	+
Dummy3	1 if the work trip took place between 8 and 9 am; 0 otherwise	+
Dummy4	1 if the work trip took place between 9 and 10 am; 0 otherwise	+
Dummy5	1 if the non-work trip took place between 9 and 10 am; 0 otherwise	+
Dummy6	1 if the non-work trip took place between 10 and 11 am; 0 otherwise	+
Dummy7	1 if the non-work trip took place between 11 am and 12 pm; 0 otherwise	+
Dummy8	1 if the non-work trip took place between 1 and 2 pm; 0 otherwise	+

***-/+ is the expected sign for the work and non-work models, respectively**

4.2.2 Modeling Techniques

This research considers four major methodological approaches to model trip generation at the micro (household) level in the study area; 1) regression modeling, 2) cross-classification, 3) discrete choice modeling, and 4) count modeling approaches. In addition, the predictive ability of these micro-models is compared against the conventional zone-based regression models. The techniques followed in each approach are documented below. The parameter estimation for the different models, except for the cross-classification approach, is performed in the NLOGIT 5.0 software.

1) Regression Modeling Approach

The ordinary least square (OLS) regression model can be used to capture the effect of different population characteristics on the number of trips generated per household/census tract in the study area. That is, given certain population characteristics, for instance age categories and information on vehicle ownership, this model predicts the number of generated trips per household/census tract *n* using the following linear formula:

$$y_n^* = \beta_1 x_{1_n} + \beta_2 x_{2_n} + \cdots + \varepsilon_n, \quad n = 1,2,\cdots,N \tag{4.1}$$

where y_n is the number of generated trips, n indexes the nth observation (which is either the household or the census tract), β is the vector of the parameters that needs to be estimated, x is the vector of independent variables, and ε is a random disturbance that follows the normal distribution with a zero mean and constant standard deviation.

2) Cross-classification Approach

The cross-classification analysis separates the population in the study area into relatively homogenous groups based on certain socio-demographic characteristics. This is followed by empirically estimating the average production rates per household for each class. Hence, creating a lookup table that can be used to forecast trip production rates at the household level depending on which class category that household belongs to. The socio-demographic characteristics used in this study are vehicle ownership (number of vehicles owned by a household), gender (number of males in the household), and household size.

3) Discrete Choice Modeling Approach (Ordered Logit)

Ordered logit models are used when the dependent variable is ordinal. In each case of this project, the numbers of generated trips y* by household are categorized as either 0, 1, 2, or 3 or more trips. Categories 0, 1, and 2 are assigned the same numerical values in the model. As for the category representing three or more trips, it is assigned a numerical value of 3 in the models. As such and from Train's (2009) discussion on trip decisions made by an individual or a household, the following applies:

- If $U_i > \mu_1$, then $y^* = 0$
- If $\mu_1 < U_i < \mu_2$, then $y^* = 1$
- If $\mu_2 < U_i < \mu_3$, then $y^* = 2$
- If $U_i < \mu_3$, then $y^* \geq 3$

The utility of household i is then determined using the following systematic and random components:

$$U_i = \beta X_i + \varepsilon_i \tag{4.2}$$

where X_i is a vector of the independent variables and β is the set of coefficients to be estimated. As for the random component, ε_i, captures the unobserved factors and errors. An ordered logit model follows the assumption that ε_i follows a gumbel distribution.

The probability of household i to generate 0, 1, 2, or 3^+ trips can be derived from the utility function. For instance, the probability of generating 0 trips is given as follows:

$$\Pr(0) = \Pr(U_i > \mu_1) = \Pr(\beta X_i + \varepsilon_i > \mu_1) = \Pr(\varepsilon_i < \mu_1 - \beta X_i) \tag{4.3}$$

As for the probability of generating 1 trip;

$$\Pr(1) = \Pr(\mu_1 < U_i < \mu_2) = \Pr(\mu_1 < \beta X_i + \varepsilon_i < \mu_2)$$

$$\Pr(1) = \Pr(\varepsilon_i < \mu_2 - \beta X_i) - \Pr(\varepsilon_i < \mu_1 - \beta X_i)$$

The probabilities for generating 2 and 3^+ trips are estimated in the same fashion. The sum of the four probabilities must add up to 1. Also, the threshold values (μ) are determined on the basis of these four categories. From a practical perspective, the predicted number of generated trips for a given household *i* can be calculated using the following equation:

$$y_i^* = 0 \cdot \Pr(0) + 1 \cdot \Pr(1) + 2 \cdot \Pr(2) + 3.2 \cdot \Pr(3^+) \tag{4.4}$$

where the 3.2 factor used in the weighted probabilities of 3 or more trips represent the average number of generated trips under the 3^+ category. Table 4-3 shows that the majority of the generated trips fall in the 0, 1, and 2 categories.

Table 4-3 Number of households (from the 5% modeling sample) falling under each ordered category for work and all other trips

Number of Generated Trips	Work Trips		All other trips	
0	3,558	35%	3,003	30%
1	2,910	29%	3,865	39%
2	3,084	30%	2,488	25%
3+	640	6%	644	6%
Total	10,192	100%	10,000	100%

4) Count Modeling Approach (Poisson Model)

Count data models are also considered in modeling trip generation as the dependent variable (trip frequency) is a positive integer variable. The most common technique employed to model count data is Poisson regression. This study considers both the Poisson and negative binomial regression models. However, since the dataset used in the modeling process does not suffer from over-dispersion, the negative binomial model was not preferred.

In the Poisson model, the probability of specific household i making 0, 1, 2, or 3^+ trips (y_i), is given by:

$$P(y_i|X_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!} \tag{4.5}$$

Where the Poisson parameter λ_i is determined by:

$$\lambda_i = \exp(y_i) = \exp(\beta X_i) \tag{4.6}$$

4.2.3 Model Validation Techniques

The comparison between the trip generation models does not only depend on the goodness-of-fit for each model, represented by the R^2 and McFadden's ρ^2, but rather on the model's ability to predict future behavior. In addition, the models are estimated and calibrated using a 5% random sample consisting of 10,000 households from the London CMA synthesized population. As a result, to validate the models' predictive ability, another 5% random sample is selected from the remaining 95% of the synthesized population. Using the 5% cross-validation sample, different validation measures are considered to check the correlation, accuracy, and coincidence ratio between the observed and predicted trip generation rates.

The first measure considered is the correlation coefficient (r) as it reflects the strength and direction of the linear relationship between two variables (i.e., observed and predicted trip rate). The correlation coefficient ranges from -1.0 to 1.0 inclusive, with values close to 1.0 suggesting strong positive relation between the predicted and observed values. The following formula can be used to calculate the correlation coefficient:

$$r = \frac{\sigma_{y_n y_n^*}}{\sigma_{y_n}\sigma_{y_n^*}} \tag{4.7}$$

where y_n^* refers to the predicted trip rate and y_n refers to the observed trip rate, σ_{y_n} and $\sigma_{y_n^*}$ are the standard deviations for y_n and y_n^*, respectively, and $\sigma_{y_n y_n^*}$ is the covariance between y_n and y_n^*.

The second and third measures considered are the root-mean-square error (RMSE) and percent RMSE (%RMSE), respectively. The RMSE is a measure of accuracy of the trip rate measuring the average error between the observed and predicted trip rate. Typically, a RMSE close to 0 suggests a strong predictive ability. Since RMSE is measured on the same scale for all the observations, a scaling problem may arise. Hence, %RMSE is also considered as it normalizes the RMSE and eliminates the scaling effect. Again, %RMSE values close to 0 are the most favorable. The formulas used to estimate RMSE and %RMSE are:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n^* - y_n)^2} \tag{4.8}$$

$$\%RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{y_n^* - y_n}{y_n}\right)^2} \tag{4.9}$$

It is insufficient to only check the correlation and accuracy of the predicted trip rates; the frequency distribution of trip rates is also important and must be checked. Therefore, the coincidence ratio (CR) is used as the fourth measure to determine the percent area that matches for the frequency (i.e., household count) of the observed and predicted trip rates categorized in four groups; 0, 1, 2, and 3+. The coincidence ratio can be calculated using the following formula:

$$CR = \frac{\sum_{TC}[\min(H_{TC}, H_{TC}^*)]}{\sum_{TC}[\max(H_{TC}, H_{TC}^*)]} \tag{4.10}$$

where H_{TC} and H_{TC}^* are the total number of households observed and predicted at trip count TR, respectively. A CR value close to 1.0 suggests a superior predictive ability of the utilized model.

4.2.4 Micro Versus Zone-Based Models

As stated in the introduction, the second objective of this study is to compare the predictive ability of the estimated trip generation micro-models against the conventional zone-based models. To estimate a zone-based model, the frequency of trips generated per household are aggregated for each census tract (i.e., TAZ). Then, socio-demographic variables also aggregated to the zonal level (namely, population belonging to certain age groups, and vehicle ownership on the TAZ level) are employed to estimate a zone-based regression model.

To compare the zone-based regression model and the micro-models, the micro-models are applied to the entire synthesized population and the trip frequency is aggregated to the zonal level. The validation methods used in this comparison are correlation coefficient, RMSE, %RMSE, and residual. The residual is simply the difference between the observed trip rate and the predicted trip rate.

4.3 Modeling Trip Distribution

This study aims to model trip distribution as a destination choice problem for shopping trips. The shopping destinations considered are based on the household's choice destinations from the LHTS. Since the decision makers are bound by their knowledge of the destinations, instead of having a universal choice set, each traveler will have a unique (constrained) choice set. Therefore, as stated in the literature review chapter, the later will reduce the bias that the traveler is knowledgeable about all the destinations in the choice set. The choice sets for the trip distribution models will be created at the micro-level in two different ways. The first way to form the choice sets consider the actual locations of the stores as the possible destinations. The second way considers the locations of the TAZs where the trips ended as the possible destinations.

4.3.1 Data Used

The households that engaged in shopping trips are identified and extracted from the 5% random sample. This results in 3,060 shopping trips (records) belonging to 1,053 households.

Each record corresponds to a home-based shopping trip undertaken by an individual. The 1,053 households and their shopping destinations are geocoded on the London road network using ArcGIS 10.1. The shopping destinations are first explored and categorized as shown in Table 4-4. The destinations that are in close proximity of each other are merged into one shopping opportunity. Therefore, the 329 original destinations are reduced to 127 shopping opportunities. In addition to the information on the destination type, each record also includes information on each traveler and the household he/she belongs to.

Table 4-4 Classification of the destinations visited in the analysis

Store Type	Number of Stores	% of Stores	Category
Grocery Stores	45	13.7%	Food
Fruit and Vegetable Markets	1	0.3%	Food
Candy Stores	1	0.3%	Food
Retail Bakeries	1	0.3%	Food
Walmart	4	1.2%	Food
Costco	2	0.6%	Food
Cigar Shops	3	0.9%	Other
Alcoholic Beverages	15	4.6%	Food
Shoppers Drug Mart	15	4.6%	Medicine
Other Pharmacies	20	6.1%	Medicine
Discount Stores	18	5.5%	Other
Department Stores	30	9.1%	Other
Variety stores	30	9.1%	Other
Hardware stores	18	5.5%	Other
Garden Supply Stores	2	0.6%	Other
Motor Vehicle Dealers	12	3.6%	Other
Auto and Home Supply Stores	14	4.3%	Other
Women's Clothing Stores	4	1.2%	Other
Shoe Stores	4	1.2%	Other
Furniture Stores	5	1.5%	Other
Sporting Goods Stores and Bicycle Shops	14	4.3%	Other
Book Stores	4	1.2%	Other
Stationery Stores	4	1.2%	Other
Jewellery Stores	3	0.9%	Other
Hobby, Toy, and Game Shops	5	1.5%	Other
Florists	9	2.7%	Other
Shopping Malls	9	2.7%	Other
Commercial Banks	37	11.2%	Other
Total	329	100%	

4.3.2 Micro-level Analysis – Location of Stores as Destinations

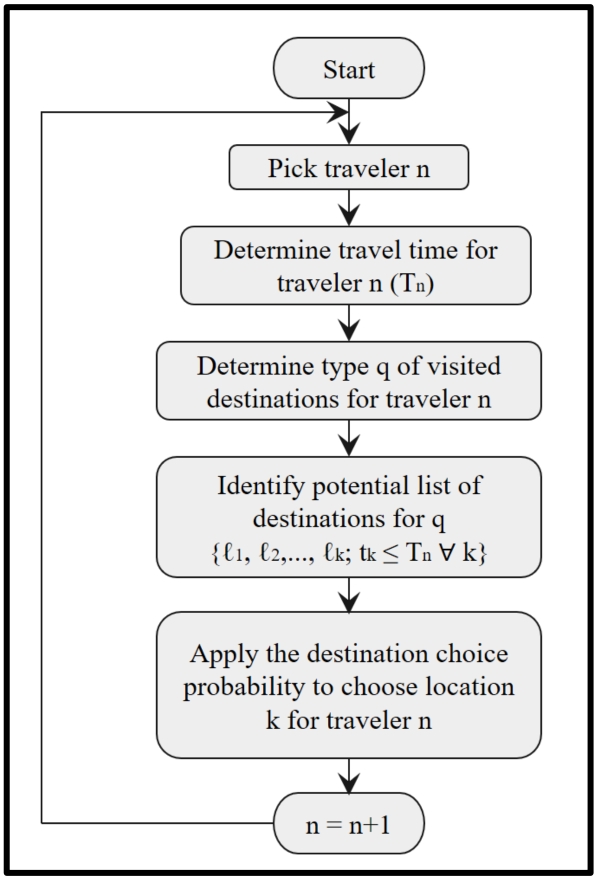

Figure 4-1 Trip distribution – Location of store as destination framework

A framework is developed to model shopping trip distribution as illustrated in Figure 4-1. This process consists of four consecutive steps: 1) estimating travel time for traveler *n* using origin and destination addresses, 2) determining the type *q* of visited destination for traveler *n* (i.e., food, medicine, or other), 3) identifying the potential list of destinations type *q* restricted by travel time for traveler *n*, and 4) applying the destination choice probability to choose location *k* for traveler *n*.

Starting with the micro-destination analysis where the destinations are the store locations, the first step is to estimate the travel time in minutes between the household residential location and his or her chosen destination for each traveling household *n*. This is done using the London road network and the Network Analyst extension in ArcGIS 10.1. The calculated travel time per individual is used to constrain the choice set for that individual. Here, we assume that the time window for the reached destination, as revealed by the observed data, is the maximum range the individual is willing to travel to his or her chosen destination. Therefore, in the absence of any other information, all the destinations within the calculate time window (i.e., maximum range) are considered as potential destinations for the traveler. Given that the traveler is using the road network, the calculated time window is used to create a service area polygon on the road network to help identify the potential destinations within it.

For the second step, the discrete choice modeling technique namely, the Multinomial Logit (MNL) model is used to predict the type of visited shopping destination. The twenty-eight shop types are further divided into three destination categories (i.e., food, medicine, and other) as shown in Table 4-3.

The third step in the micro-level framework is to identify the potential list of destinations for each traveler. We assume that the traveler's choice set is dependent on the travel time and destination type identified in the first two steps. Finally, using the constrained choice sets as the fourth step, the MNL technique is used to predict the destination *k* for each traveler *n*.

4.3.2.1 Destination Type Choice Modeling

Discrete choice models can be used to analyze and predict the traveler's choice of one destination type from a finite set of alternatives. This research makes use of the MNL models to determine the destination type visited by each traveler, from three destination types, as shown in

Figure 4-2. The estimation of the parameters for the different models are performed in the NLOGIT 5.0 software (Greene, 2011).

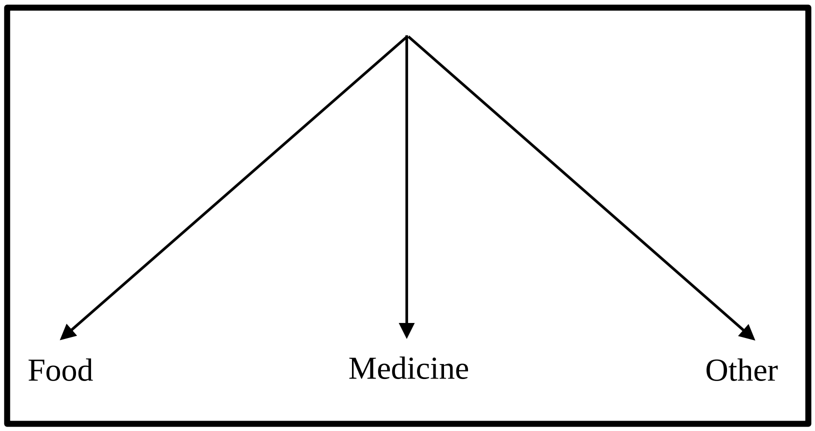

Figure 4-2 Discrete choice model structure

Each traveler *n* is associated with a utility function, U_{nk} that can be expressed as follows:

$$U_{nk} = \beta X_{nk} + \varepsilon_{nk} \tag{4.11}$$

where X_{nk} is a vector of the independent variables and β is the set of coefficients to be estimated. As for the random error term, ε_{nk} (for all *n* and *k*) are assumed to be independently and identically distributed (iid) across alternatives and observations following a Gumbel probability density function. Accordingly, the type of a given destination *k* is modelled by calculating the probability that traveler *n* will visit a specific destination type *k* such that:

$$P_{nk} = \Pr(U_{nk} > U_{nq}) = \Pr(\beta X_{nk} + \varepsilon_{nk} > \beta X_{nq} + \varepsilon_{nq}) \; for\ all\ k \neq q\ and\ n$$

The MNL choice probability can be represented as follows:

$$P_{nk} = \frac{\exp(\beta X_{nk})}{\sum_{q=0}^{J} \exp(\beta X_{nq})} \tag{4.12}$$

Model Formulation

Two types of destination attributes are considered in the shopping destination type model in this research: shopping destination categories (Table 4-4) and temporal factors. Twenty-eight dummy variables corresponding to the shopping categories are created to help determine the destination type (i.e., food, medicine, or other). In addition, a store diversity index is estimated on the zonal level to indicate the store variety in the zone. Since there are twenty-eight different store types in the study area, equation 4.13 is used. The results of the diversity index for 55 TAZs are illustrated in Figure 4-3.

$$Diversity = \frac{Num.of\ store\ types\ in\ a\ zone}{28} \tag{4.13}$$

As for the temporal factors, these include the maximum range of the service area (i.e., traveler's trip time window), the distance from the shopping destination to the nearest Highway and the distance from the shopping destination to the Central Business District (CBD). These factors represent the accessibility of each destination.

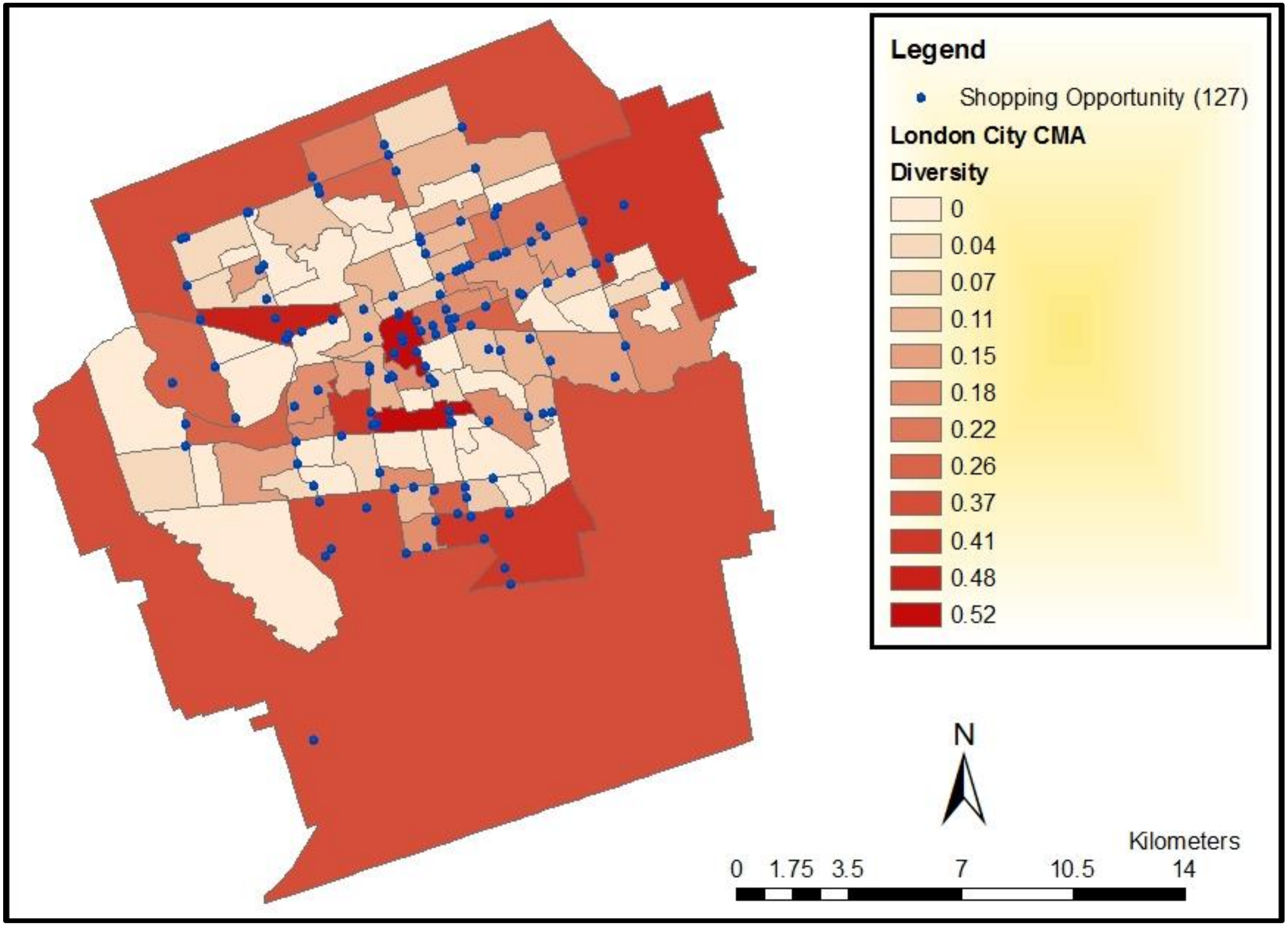

Figure 4-3 The results of the diversity index for the London TAZ

4.3.2.2 Constrained and Unconstrained Destination Choice Sets

For the unconstrained destination choice set, all 127 shopping opportunities are considered in the choice set of each traveler *n* in the MNL model. As a result, we assume that all of the shopping opportunities are accessible to the travelers. On the other hand, the constrained destination choice set is unique for each traveler based on two factors; the estimated time window and destination type. The estimated time window for each individual in the first step is considered as the time budget the individual is willing to travel to his/her destination. As noted earlier, a polygon is then created for each traveler to represent the area that can be reached within a specified amount of time (i.e., calculated in the first step). This polygon determines the range of the traveler's service area and is used to identify the shopping opportunities that are within that service area. For example, as shown in Figure 4-4, an individual in household 823 reached his/her destination (i.e., ID 115) in 6 minutes; therefore, a 6-minute service area is created. Consequently, the shopping opportunities are reduced from the universal set of 127 destinations to a set of 31 constrained choices that fall within the traveler's 6 minutes service area, including the actual choice.

Moreover, using the destination type model developed in the second step, the type of destination for the traveler is predicted. Using the same example as above, the individual's predicted destination type is food. Therefore, only the shopping opportunities pertaining to food stores are left in the individual's constrained choice set as shown in Figures 4-5. Hence, the constrained choice set is further reduced to 14 shopping opportunities.

Table 4-5 shows the frequency of the number of alternatives in the constrained choice set with the average number of alternatives being 21. The large variance indicates that the travelers are faced by different choice sets. For the unconstrained choice set, all the destinations are assumed to be accessible to the traveler and all the choice sets will include 127 alternatives (i.e., shopping opportunities).

Table 4-5 Descriptive statistics of alternatives in the constrained choice sets

Number of alternatives in constrained choice set			Frequency (% of total)
	1-10		910 (29%)
	11-20		677 (21%)
	21-30		789 (25%)
	31-40		390 (12%)
	41-50		394 (13%)
	Total		3,160 (100%)
Sample size	**mean alternatives**	**min alternatives**	**max alternatives**
3,160	21	2	47

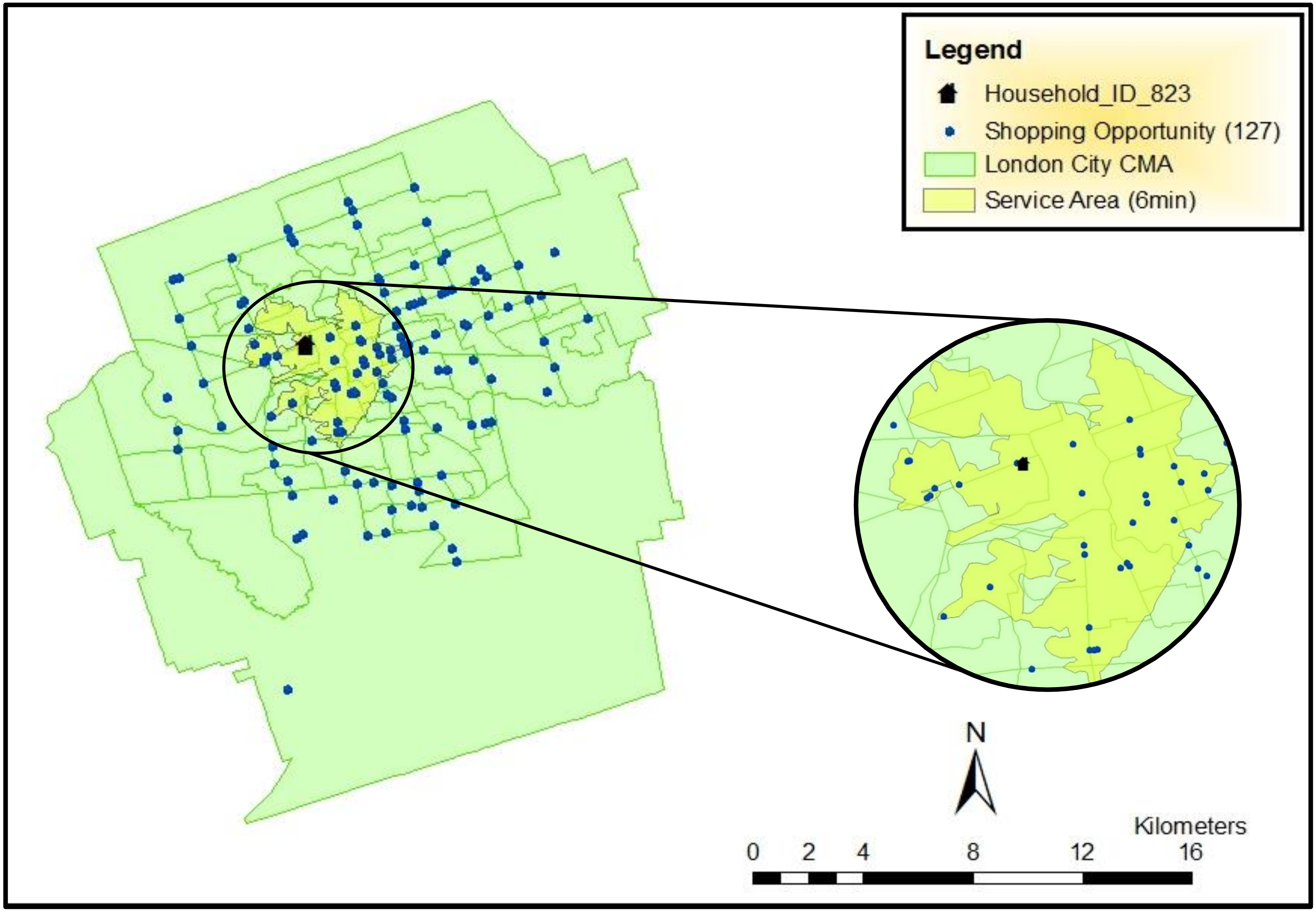

Figure 4-4 Example on using service area to constrain the destination choice set

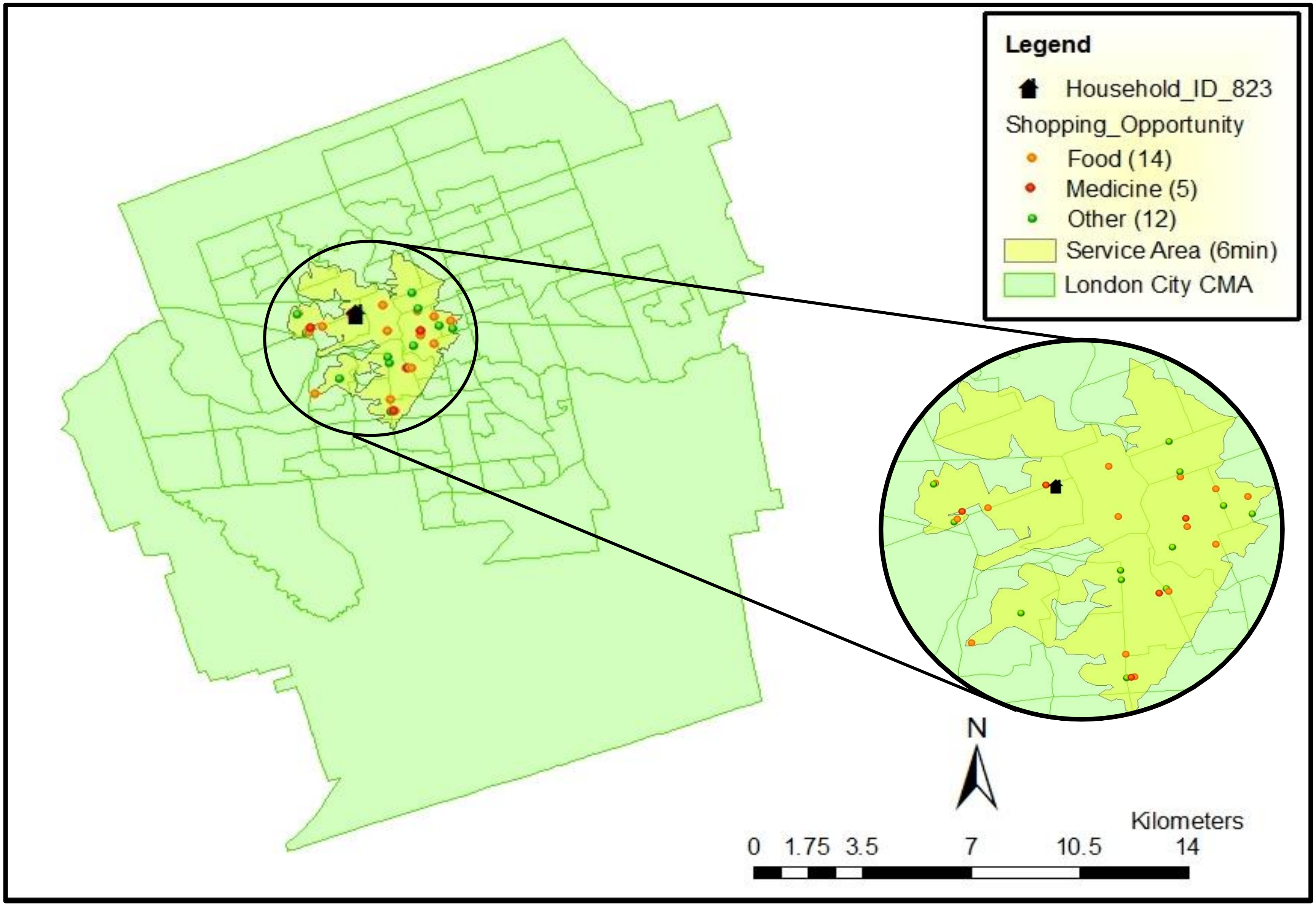

Figure 4-5 Example on using destination type to constrain the destination choice set

4.3.2.3 Destination Choice Modeling

In this part of the research, the MNL modeling technique is also used to determine the destination *k* for traveler *n*. The choice model structure is represented in Figure 4-6. The specifications for the MNL model are illustrated in section *4.3.2.1*.

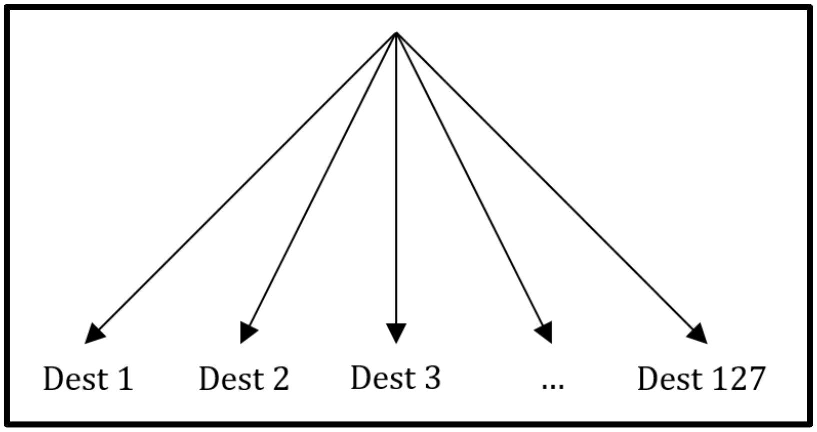

Figure 4-6 Destination choice model structure – Unconstrained Case

Model Formulation

The independent variables considered in the destination choice modeling include the variables used in the destination type modeling. Aside from the shopping destinations and temporal variables, Table 4-6 lists other socio-economic and demographic characteristics for the travelers and the households in the model. Three models are estimated for the constrained and unconstrained choice sets, each of which is based on different combinations of explanatory variables.

Table 4-6 List of independent variables considered in destination choice model

Variable	Description
Age (<20)	1 if traveler belongs to the < 20 age group; 0 otherwise
Age (20-34)	1 if traveler belongs to the 20-34 age group; 0 otherwise
Age (35-49)	1 if traveler belongs to the 35-49 age group; 0 otherwise
Age (50-64)	1 if traveler belongs to the 50-64 age group; 0 otherwise
Age (65$^+$)	1 if traveler belongs to the 65$^+$ age group; 0 otherwise
Employed	1 if the traveler is employed (full-time and part-time); 0 otherwise
Female	1 if the traveler is a female; 0 otherwise
Household Size	The number of people living in the traveler's household
Walked	1 if the traveler walked to the destination; 0 otherwise

4.3.3 Micro-level Analysis – Location of TAZs as Destinations

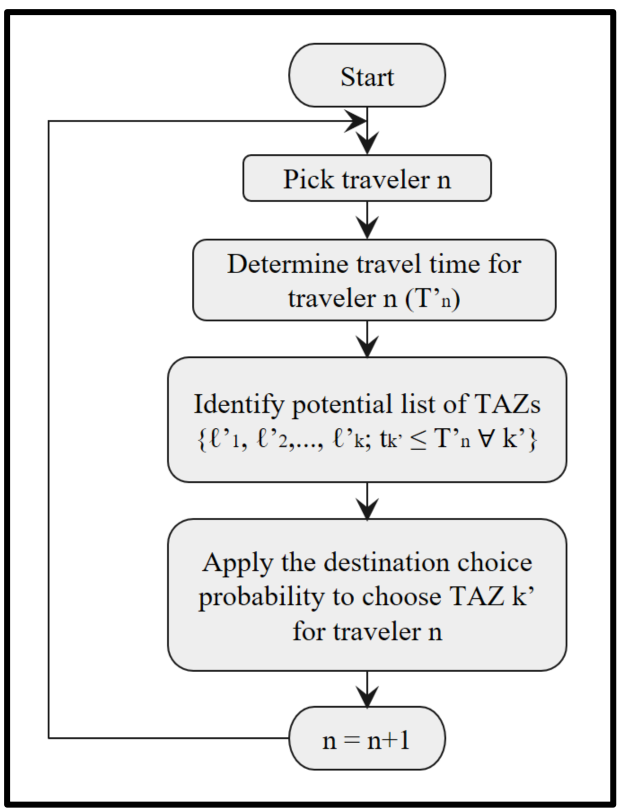

Figure 4-7 Trip distribution micro-level analysis – location of TAZ as destination framework

Figure 4-7 summarizes the framework followed to model shopping trip distribution by using the Traffic Analysis Zones (TAZs) as shopping destinations. The process consists of three main steps: 1) estimating travel time for traveler *n* from the household to the centroid of the chosen zone, 2) identifying the potential list of zones restricted by travel time for traveler *n*, and 3) applying the destination choice probability to choose destination zone *k* for traveler *n*.

Before beginning the micro-level analysis where the destinations are the TAZs, the shopping opportunities are aggregated to the TAZ level. This results in fifty-five alternative TAZs to form the unconstrained alternative zones (i.e., universal choice set), as the other TAZs do not have shopping facilities (see Figure 4-8). The same steps used in the micro-level analysis where the destinations are the store locations are also followed in this section. The first step is to estimate the travel time in minutes between the traveler's household residence and the centroid of the observed chosen zone for each traveler *n*. This is also done using the London road network and the Network Analyst extension in ArcGIS 10.1.

For the second step, the potential list of TAZs for each traveler is constrained based on time window calculated in the previous step. An example is provided in Figure 4-9, in which the constrained list of TAZs for the traveler in household ID 823 is reduced from the universal choice set of 55 alternatives to the constrained choice set of 13 TAZs. Finally, in the third step, the constrained choice sets and the MNL technique are used to model the destination TAZ for each traveler.

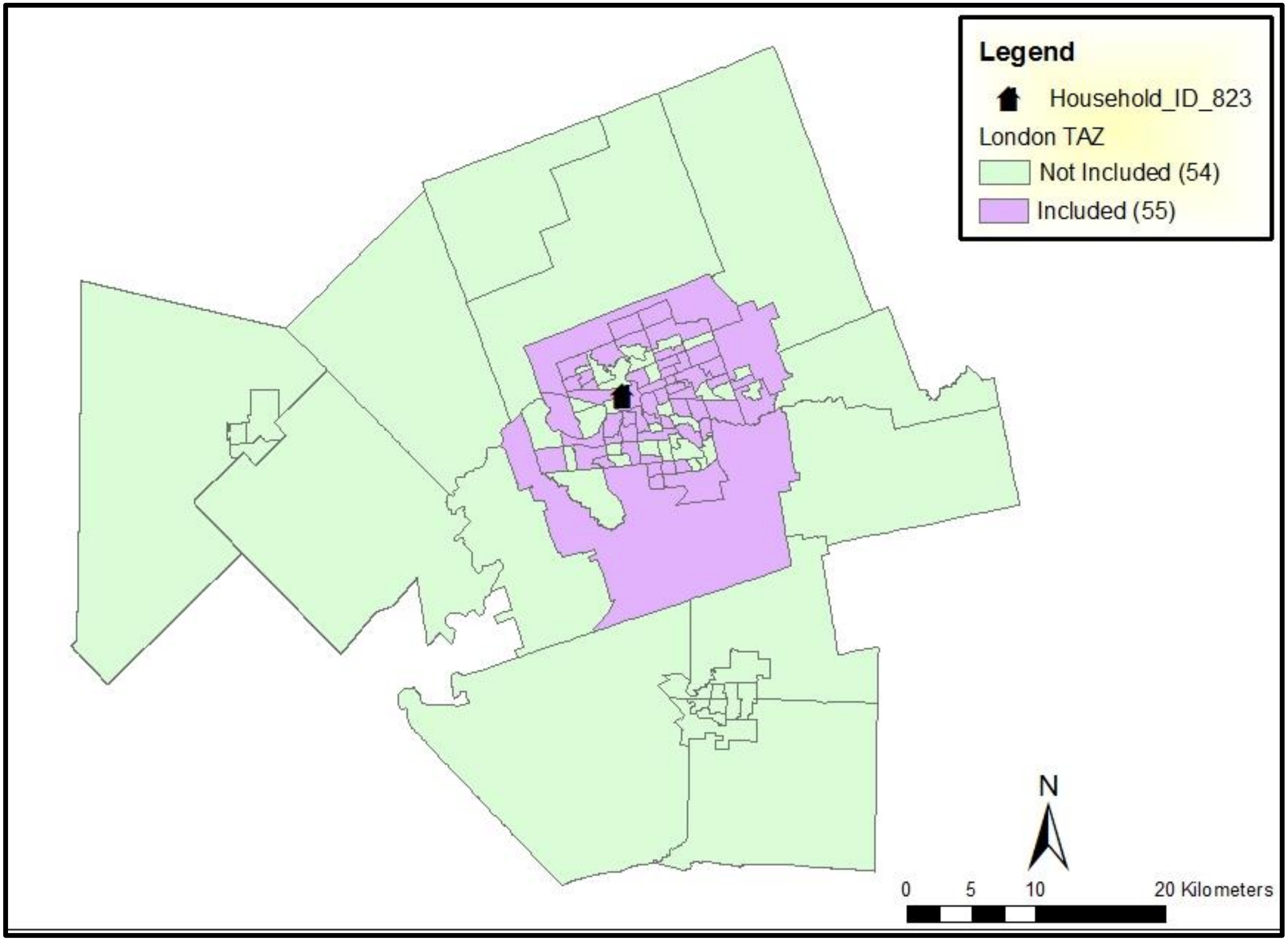

Figure 4-8 Example on TAZ unconstrained choice set

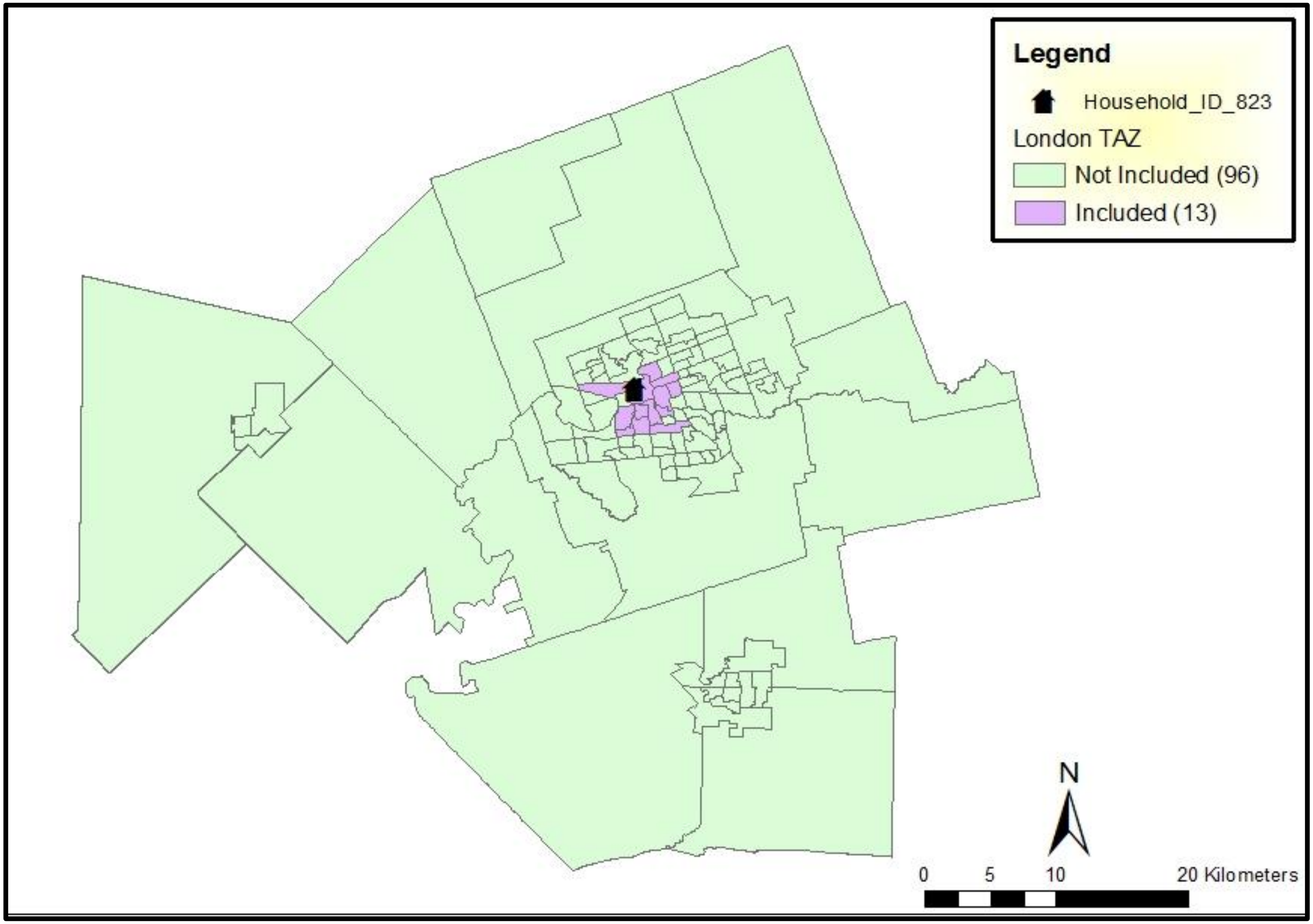

Figure 4-9 Example on TAZ constrained choice set

4.3.3.1 Constrained and Unconstrained Destination Choice Sets

While the unconstrained choice set consisted of 55 alternative destinations, the constrained choice sets varied for the 3,160 travelers, as shown in Table 4-7. The table provides the frequency of the number of TAZ alternatives in the constrained choice set with the average number of TAZs being 14. Therefore, even when considering the TAZs instead of the actual location of the shopping opportunities, there is a large variance suggesting that travelers are faced with different choice sets.

Table 4-7 Descriptive statistics of TAZs in the constrained choice sets

Number of TAZs in constrained choice set			Frequency (% of total)
1-5			582 (18%)
5-10			634 (20%)
11-15			471 (15%)
16-20			770 (24%)
21-25			518 (17%)
26-30			185 (6%)
Total			3,160 (100%)
Sample size	**mean alternatives**	**min alternatives**	**max alternatives**
3,160	14	2	29

4.3.4 Model Validation Techniques

The comparison between the shopping trip distribution models is not only dependent on the goodness-of-fit for each model, represented by McFadden's ρ^2. It also depends on the model's ability to predict the traveler's destination. As stated earlier, the models are estimated and calibrated using the 5% random sample consisting of 10,000 households from the London CMA synthesized population. Therefore, the same 5% random sample used in the validation of the trip generation models is used to validate the trip distribution models' predictive ability.

Using the 5% cross-validation sample, the following measures are considered to determine the model to be utilized; percent predicted right and margin of error. The margin of error checks the travel time in minutes between the predicted and actual shopping destination.

CHAPTER 5: RESULTS AND DISCUSSION

5.1 Modeling Trip Generation

5.1.1 Micro-Analysis

The results of the trip generation micro-models for work trips are summarized in tables 5-1 and 5-2. The models show that work trip generation is effected by the household structure (i.e., age categories and gender), vehicle ownership, and the hour of the day the trip took place. For the models presented in Table 5-1, the relationship between age and work trip generation is non-linear. Taking Age1 (<20 years) as the reference category, the results show that work trip frequency per household increases as age increases, reaching a peak for the working age group (Age3), and then it decreases for the senior age group (Age 5). As expected, males, vehicle ownership, and AM-peak period have a positive impact on work trip frequency. These findings are consistent across all the micro-trip generation models for work trips. The regression model's explanatory power of work trip generation is at an empirically acceptable level with an R-squared of 0.739. Similarly, the ordered logit and Poisson models have empirically acceptable explanatory powers with McFadden's rho-squared values of 0.421 and 0.238, respectively. As for the cross-classification model (see Table 5-2), the values are presented as work trip rate per household, number of males, and vehicle ownership. The results follow the findings in the previous models as work trip rate increases with increase of males and vehicle ownership in a household.

As for the results of the trip generation micro-models for non-work trips, they are seen in tables 5-4 and 5-5. All the coefficients in the model estimates have the expected signs. Similar to the work trip micro models, non-work trip generation is also affected by the household structure (i.e., age categories and gender) and the hour of the day the trip took place. For the three models presented in Table 5-4, the relationship between age and non-work trip generation is non-linear. Holding Age1 (<20 years) as the reference category, the results vary slightly between the three models for Age2 and Age3, but for the most part they show that non-work trip rate increases as age increases, reaching a peak for the senior age group (Age5). As for using the hours of the day the trip took place in as dummy variables, they proved to be positive and highly significant for the periods between 9 am to 12 pm, and 1 pm to 2 pm. In addition, the Social Dummy is positive and highly significant, thus compared to shopping trips, social trips are more likely to increase the probability of a non-work trip being generated per household. The interaction term Social X Females also followed our expectations as females are more likely to generate social trips compared to males. The results show a satisfactory goodness-of-fit for the ordered and Poisson models with McFadden's rho-squared values of 0.414 and 0.246, respectively. However, the R-squared value for the regression model is relatively low (0.612). Finally, for the cross-classification model (see Table 5-5), the values are presented as non-work trip rate per household, number of females, and vehicle ownership. The results vary considerably for the different household structures and vehicle ownership, but no clear pattern is observed.

5.1.2 Validation Results for Micro-analysis

The performance comparison between the micro-models is conducted based on out-of-sample validation. As stated in the method of analysis section, the measures of correlation, variance (RMSE and %RMSE), and coincidence are utilized for the validation assessment. Tables 5-3 and 5-6 summarize the results of validation for work and non-work trips, respectively. For the work trip micro-models, the results do not vary significantly for the first two models.

However, the four considered measures are in favor of the ordered logit model, followed by the regression, Poisson, and cross-classification models, respectively.

On the other hand, the validation results for the non-work trip micro-models vary significantly for the cross-classification model. The later model performs very poorly proving that using household size, females, and vehicle ownership on their own is not very useful when predicting non-work trips in a household. As for the three other models, the validation results are similar and comparable. The ordered logit model has the most accurate predictions followed by the regression and Poisson models, respectively.

Table 5-1 Results from trip generation models for work trips

Variables	Regression			Ordered Logit			Poisson			Negative Binomial		
	Coeff	Std Err	t-stat	Coeff	Std Err	t-stat	Coeff	Std Err	t-stat	Coeff	Std Err	t-stat
Constant	-0.01	0.013	-0.94	-2.76	0.076	-36.58	-1.00	0.030	-32.65	-1.00	0.048	-20.62
Age2	0.28	0.009	29.28	1.23	0.049	25.19	0.21	0.017	12.64	0.21	0.027	7.89
Age3	0.34	0.012	29.28	1.41	0.058	24.28	0.29	0.022	13.47	0.29	0.036	8.13
Age4	0.22	0.010	21.14	0.98	0.052	18.69	0.20	0.020	9.78	0.20	0.032	6.12
Age5	-0.033	0.011	-2.96	-0.36	0.065	-5.56	-0.62	0.035	-17.48	-0.62	0.038	-16.43
Males	0.03	0.009	3.10	0.14	0.043	3.21	0.03	0.016	1.57	0.03	0.028	0.91
Vehicles	0.09	0.007	11.89	0.39	0.037	10.42	0.10	0.014	7.37	0.10	0.019	5.44
Dummy1 (6am)	0.66	0.016	42.40	3.01	0.082	36.58	0.47	0.026	17.88	0.47	0.046	10.07
Dummy2 (7am)	0.70	0.012	57.15	3.12	0.069	45.57	0.54	0.022	24.86	0.54	0.040	13.38
Dummy3 (8am)	0.70	0.012	57.58	3.13	0.068	45.94	0.53	0.022	24.46	0.53	0.040	13.30
Dummy4 (9am)	0.63	0.016	40.05	2.92	0.083	35.16	0.46	0.026	17.59	0.46	0.048	9.57
Mu(01)	-	-	-	3.61	0.055	65.13	-	-	-	-	-	-
Mu(02)	-	-	-	8.73	0.106	82.71	-	-	-	-	-	-
Alpha	-	-	-	-	-	-	-	-	-	0.0001	0.001	0.11
Sample Size	10,000			10,000			10,000			10,000		
Estimator	OLS			MLE			MLE			MLE		
F-statistics	2,825.47			-			-			-		
$-2(L(0)-L(\beta'))$	-			12,450.09			6,159.24			1.80		
R^2	0.739			-			-			-		
ρ^2	-			0.421			0.238			-		

Table 5-2 Results from trip generation models for work trips (Cross-Classification models)

Household Size	Males	Vehicle Ownership				Household Size	Males	Vehicle Ownership			
		0	**1**	**2**	**3+**			**0**	**1**	**2**	**3+**
1	**0**	0.11	0.21	0.00	0.00	**4**	**0**	0.58	0.24	1.14	0.43
	1	0.30	0.47	0.63	0.00		**1**	0.42	1.01	1.35	1.56
	2	0.00	0.00	0.00	0.00		**2**	0.23	1.53	1.57	1.76
	3	0.00	0.00	0.00	0.00		**3**	0.95	1.14	1.67	1.68
	4+	0.00	0.00	0.00	0.00		**4+**	0.00	2.18	2.26	0.00
2	**0**	0.11	0.60	0.85	0.00	**5**	**0**	0.00	0.00	1.52	0.00
	1	0.54	0.46	0.93	0.64		**1**	0.00	0.98	1.13	1.21
	2	0.72	0.83	1.07	1.56		**2**	0.00	0.74	1.20	1.88
	3	0.00	0.00	0.00	0.00		**3**	0.00	1.38	1.79	1.64
	4+	0.00	0.00	0.00	0.00		**4+**	0.00	3.67	1.23	2.00
3	**0**	0.35	0.72	0.61	0.53	**6+**	**0**	0.00	0.00	0.00	0.00
	1	0.91	1.08	1.42	1.50		**1**	0.00	0.95	0.71	1.05
	2	0.19	1.09	1.51	1.64		**2**	1.53	0.82	1.16	1.96
	3	0.00	1.48	1.35	1.23		**3**	0.00	4.02	1.81	1.73
	4+	0.00	0.00	0.00	0.00		**4+**	0.00	0.86	2.53	2.83

Table 5-3 Validation of work trip generation models

Validation	Regression	Ordered Logit	Poisson	Cross-Classification
Correlation	0.831	0.840	0.765	0.607
RMSE	0.519	0.504	0.598	0.701
%RMSE	0.243	0.263	0.255	0.341
Coincidence Ratio	0.647	0.675	0.591	0.538

Table 5-4 Results from trip generation models for non-work trips

Variables	Regression			Ordered Logit			Poisson			Negative Binomial		
	Coeff	Std Err	t-stat	Coeff	Std Err	t-stat	Coeff	Std Err	t-stat	Coeff	Std Err	t-stat
Constant	0.05	0.012	3.72	-2.28	0.075	-30.58	-1.52	0.043	-35.35	-1.52	0.056	-26.87
Age2	0.02	0.007	2.42	-0.05	0.043	-1.11	-0.15	0.026	-5.83	-0.15	0.028	-5.45
Age3	0.03	0.007	3.96	0.03	0.046	0.70	-0.13	0.030	-4.20	-0.13	0.034	-3.73
Age4	0.15	0.008	19.21	0.70	0.043	16.21	0.33	0.025	13.08	0.33	0.031	10.72
Age5	0.356	0.010	36.47	1.47	0.054	27.14	0.51	0.027	19.13	0.51	0.032	15.75
Dummy5 (9am)	0.66	0.025	26.64	2.65	0.124	21.39	0.65	0.048	13.59	0.65	0.057	11.44
Dummy6 (10am)	0.73	0.016	46.22	2.99	0.081	36.77	0.86	0.034	25.36	0.86	0.041	20.98
Dummy7 (11am)	0.69	0.021	32.57	2.86	0.104	27.36	0.79	0.044	17.94	0.79	0.055	14.46
Dummy8 (1pm)	0.76	0.020	37.35	3.12	0.101	30.79	0.83	0.041	20.08	0.83	0.050	16.60
Social Dummy	0.46	0.023	19.63	1.84	0.113	16.37	0.38	0.053	7.19	0.38	0.074	5.18
Social × Female	0.12	0.015	8.19	0.51	0.073	7.04	0.20	0.033	6.19	0.20	0.042	4.81
Mu(01)	-	-	-	3.43	0.063	54.78	-	-	-	-	-	-
Mu(02)	-	-	-	8.33	0.171	48.60	-	-	-	-	-	-
Alpha	-	-	-	-	-	-	-	-	-	0.0001	0.01	0.01
Sample Size	10,000			10,000			10,000			10,000		
Estimator	OLS			MLE			MLE			MLE		
F-statistics	1,576.40			-			-			-		
$-2(L(0)-L(\beta'))$	-			7,622.33			4,581.87			-0.30		
R^2	0.612			-			-			-		
ρ^2	-			0.414			0.246			-		

Table 5-5 Results from trip generation models for non-work trips (Cross-Classification models)

Household Size	Females	Vehicle Ownership				Household Size	Females	Vehicle Ownership			
		0	1	2	3+			0	1	2	3+
1	0	0.37	0.26	0.37	0.54	4	0	0.00	0.00	0.27	0.00
	1	0.67	0.42	0.41	0.00		1	0.38	0.32	0.15	0.14
	2	0.00	0.00	0.00	0.00		2	0.79	0.49	0.14	0.28
	3	0.00	0.00	0.00	0.00		3	0.00	0.49	0.16	0.13
	4+	0.00	0.00	0.00	0.00		4+	0.00	0.00	0.00	0.00
2	0	0.28	0.26	0.13	0.21	5	0	0.00	0.00	0.00	0.00
	1	0.64	0.72	0.52	0.66		1	2.00	0.29	0.30	0.30
	2	0.23	0.26	0.33	0.00		2	1.00	0.77	0.59	0.22
	3	0.00	0.00	0.00	0.00		3	0.00	1.69	0.33	0.23
	4+	0.00	0.00	0.00	0.00		4+	0.00	0.00	0.34	0.00
3	0	0.00	0.05	0.10	0.27	6+	0	0.00	0.00	0.00	0.00
	1	0.83	0.24	0.16	0.25		1	0.00	1.05	0.58	0.30
	2	0.60	0.41	0.20	0.28		2	1.13	1.01	0.97	0.63
	3	0.00	0.30	0.00	0.00		3	0.00	0.22	0.58	0.73
	4+	0.00	0.00	0.00	0.00		4+	0.00	0.00	0.32	0.00

Table 5- 6 Validation of non-work trip generation models

Validation	Regression	Ordered Logit	Poisson	Cross-Classification
Correlation	0.766	0.766	0.639	0.300
RMSE	0.435	0.434	0.538	0.647
%RMSE	0.267	0.279	0.320	0.349
Coincidence Ratio	0.752	0.807	0.755	0.568

5.1.3 Zonal vs. Micro-models

Table 5-7 summarizes the results of the zone-based ordinary least squares (OLS) regression model for work trips. The coefficients in the model estimates have the expected signs. Holding *Age1* (<20 years) as the reference category, the results show that work trip frequency per zone increases as age increases, reaching a peak for *Age2* and *Age3*, and then it decreases for the senior age group (*Age 5*). In addition, vehicle ownership has a positive effect on work trip generation per zone, as expected. As for the goodness-of-fit represented by the R-squared value (0.649), it is acceptable but not astounding.

Finally, the comparison results between the zonal-model and micro-models for work trips are represented in Table 5-8 and Figure 5-1. By comparing the RMSE and %RMSE validation measures (see Table 5-8), the results of the regression, ordered logit, and Poisson micro-models are similar to the zonal regression model. While the cross-classification model's results are very poor with more than double the RMSE and %RMSE compared to the other models. As for the residual results (Figure 5-1), the micro-models with the exception of the cross-classification model are consistent across the 109 different traffic analysis zones. In conclusion, the micro-models prove to have an advantage over the zone-based model as they are able to predict work trip rates on the micro and macro level accurately.

Table 5-7 Results for work trips regression zonal model

Variables	Coeff	Error	t-stat
Age2	0.47	0.029	16.09
Age3	0.44	0.094	4.69
Age4	0.22	0.100	2.19
Age5	-0.22	0.083	-2.62
Vehicles	0.23	0.082	2.81
Sample Size	109		
Estimator	OLS		
R^2	0.649		

Table 5-8 Validation of work trip generation models (zonal vs. household)

Validation	Regression	Ordered	Poisson	Cross-Classification	Regression (Zonal)
Correlation	0.999	0.999	0.999	0.993	0.999
RMSE	30.90	30.86	26.77	72.60	28.26
%RMSE	0.024	0.024	0.021	0.052	0.024

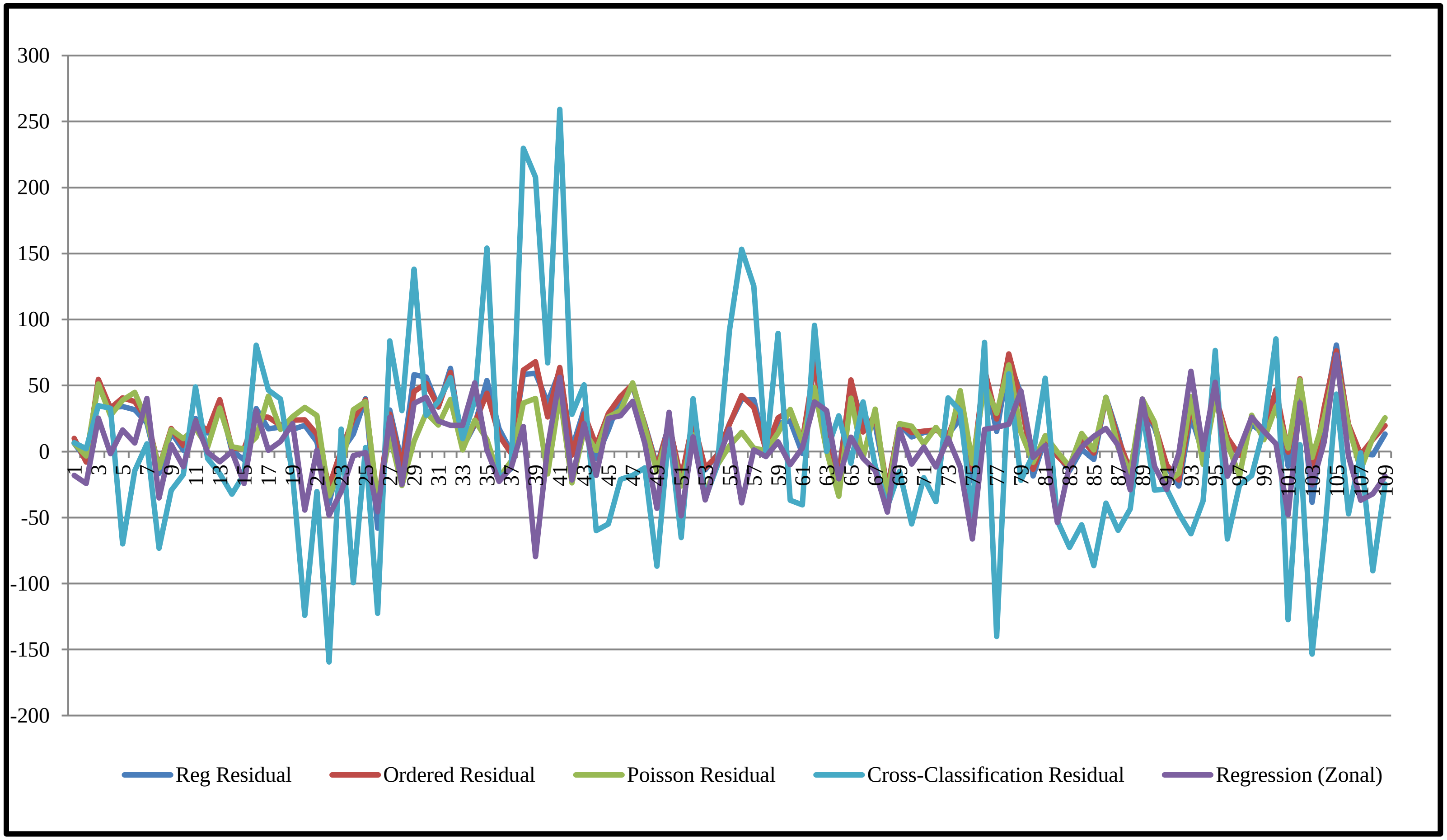

Figure 5-1 Residual plot for zonal vs. household work trip generation models

Moving to the results of the zone-based OLS regression model for non-work trips, they are shown in Table 5-9. The coefficients in the model estimates follow the expected hypotheses. Keeping the *Age1* (<20 years) variable as the reference category, the results show that non-work trip frequency per zone increases as age increases, reaching a peak for *Age3* and then decreases for the senior age group (*Age 5*). As for vehicle ownership, it has a positive effect on non-work trip generation per zone, as expected. These results, for the main part, are also in line with the results of the zone-based OLS regression model for work trips. The goodness-of-fit for the non-work trip generation model improved slightly compared to the work trip generation model from 0.649 to 0.699.

The comparison results between the zonal-model and micro-models for non-work trips are represented in Table 5-10 and Figure 5-2. By comparing the RMSE and %RMSE validation measures (see Table 5-10), the results of the regression, ordered logit, and Poisson micro-models are higher than the values in the work trip models. On the other hand, the RMSE and %RMSE values are lower than the zonal regression model. As for the cross-classification model, its results are very poor, as the case with the work trip model, compared to the four other models. Also, considering the results from the residual values illustrated in Figure 5-2, the regression micro- and zonal-models are consistent across the 109 different traffic analysis zones. However, the ordered and Poisson models show large residual values with the Poisson model having the highest residuals. In conclusion, the regression micro- and zonal-models prove to have an advantage over the other micro-models in the case of non-work trips, as they are able to predict non-work trip rates on the macro level accurately.

Table 5-9 Results for non-work trips regression zonal model

Variables	Coeff.	Error	t-stat
Age2	0.41	0.01	43.40
Age3	0.51	0.01	47.10
Age4	0.27	0.01	24.11
Age5	-0.07	0.01	-6.51
Vehicles	0.17	0.01	17.58
Sample Size	109		
Estimator	OLS		
F-statistics	4602.44		
R^2	0.699		

Table 5-10 Validation of non-work trip generation models (zonal vs. household)

Validation	Regression	Ordered	Poisson	Cross-Classification	Regression (Zonal)
Correlation	0.992	0.992	0.988	0.995	0.995
RMSE	34.71	47.03	50.10	136.50	98.39
%RMSE	0.061	0.078	0.080	0.20	0.15

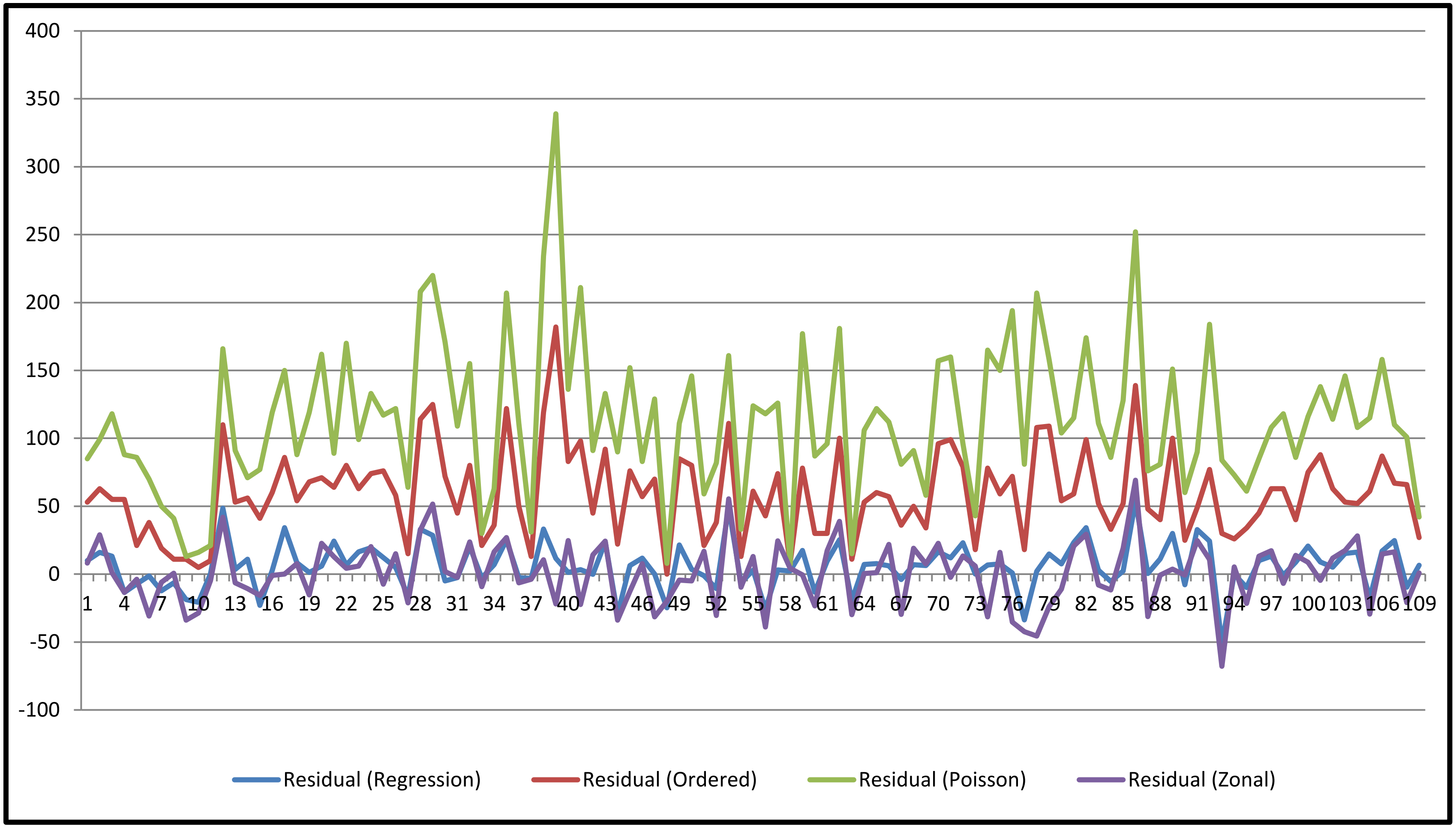

Figure 5-2 Residual plot for zonal vs. household non-work trip generation models

5.2 Modeling Trip Distribution

5.2.1 Micro Analysis – Location of Stores as Destinations

5.2.1.1 Destination Type Choice Model

A MNL model is estimated to determine the destination type an individual is visiting and the results are summarized in Table 5-11. In this model, the traveler's choice of destination type is modeled as a function of the characteristics of the alternatives and socio-economic and demographic characteristics of the decision maker. The alternative specific constants for the food and medicine type destinations are considered to represent the average effect of all the factors that influence the choice of the alternatives, but not included in the specification. The results also show that the probability of visiting a food type destination increases when the alternative is a *Grocery Store* and decreases when the destination is *Walmart*. On the other hand, the probability of visiting a medicine type destination increases when the alternative is *Shoppers Drug Mart*. Finally, the probability of visiting other type of destinations decreases when the alternative is a *Department Store*.

Another characteristic of the alternatives considered in the three utility equations is the natural logarithm of the alternative's distance to the Highway "*ln(Dist. To HWY)"*, and the results mainly show that as the destination's distance to the Highway increases, the probability of the visited destination belonging to the medicine category increases. As for the traveler's socio-economic and demographic characteristics, females of age 35-49 years (*Age3 x Female*) are more likely to visit a destination belonging to the food category. Nevertheless, a traveler of age 20-34 years (*Age2*) is more likely to visit a medicine type destination. An interaction term between travelers with a valid driving license and grocery stores (*VDL x Grocery Store*) show that they slightly decrease the probability of visiting a food type alternative. All the variables considered in the model are highly significant according to the t-stat values included in the Table. The results also illustrate a high rho-squared (ρ^2) value of 0.64 representing an acceptable goodness-of-fit. Finally, the models' out-of-sample validation shows that more than 88% of the traveler's destination type is predicted right.

Table 5-11 Results for the destination type model

Variable	Type	Coeff	Error	t-stat
$Const_1$	Food	-4.06	0.20	-20.66
Grocery Store	Food	5.05	0.25	20.31
Walmart	Food	-1.16	0.23	-5.10
ln(Dist. to HWY)	Food	-0.97	0.20	-4.85
Age3 × Female	Food	0.68	0.21	3.31
VDL × Grocery Store	Food	-0.55	0.17	-3.29
$Const_2$	Medicine	-6.08	0.29	-20.64
Shoppers Drug Mart	Medicine	8.22	0.31	26.10
ln(Dist. to HWY)	Medicine	1.26	0.19	6.58
Age2	Medicine	0.83	0.40	2.07
Department Store	Other	-2.49	0.15	-16.21
ln(Dist. to HWY)	Other	-1.22	0.08	-14.46
Number of Observations	3,160			
Log Likelihood Function (β)	-1,163.74			
Log Likelihood Function (0)	-3,215.46			
ρ^2	0.64			
% Predicted Right	88.30%			

5.2.1.2 Destination Choice Models

Initially, a MNL model is estimated based only on the 1,946 shopping trips available from the 2009 London Household Travel Survey (see Table 5-12). Next, a MNL model is estimated using the same specifications but with the 5% sample derived from the synthetic household list. Figure 5-3 compares the results for the survey and the

5% sample models. The results are very similar and consistent suggesting that the synthesized population could be used to model the destination choice problem. The use of the 5% random sample enables us to validate the estimated models using out of sample observations.

Table 5-13 and Figure 5-4 compares the unconstrained and constrained models for the shopping destination choice models using the same specifications. Although some of the variables are statistically insignificant, they are kept for comparison purposes. For the main part, the unconstrained and constrained models are similar. However, notable differences exist for the coefficients of the *Maximum Travel Window* variable and the interaction term between travelers younger than 20 years and Shoppers Drug Mart (*Age (<20) × Shoppers Drug Mart*). The *Range of Service Area* in the unconstrained model is highly insignificant. On the other hand, the *Range of Service Area* is positive and highly significant in the constrained model. The latter result is expected given the method used to restrict the destination choice set. In the two models, the interaction term (*Age (<20) × Shoppers Drug Mart*) reduces the attractiveness of destination utility if the destination choice is Shoppers Drug Mart. The utility decreases more in the unconstrained choice set model compared to the constrained one.

The coefficient of each store category reveals the contribution of each specific store type in the destination utility. Most store categories increase the utility functions, while *Other Pharmacies* category reduces the destination attractiveness. In addition, the large magnitudes of *Walmart* and *Shopping Malls* demonstrate their influential roles as shopping destinations. The coefficients for the natural logarithm of the destination's distance to the Highway (*ln(Dist. To HWY)*), are significant and negative for the unconstrained and constrained models. This indicates that the greater the distance between the destination and Highway, the smaller the propensity that an individual is likely to choose the destination. On the contrary, the coefficients for the natural logarithm of the destination's distance to the CBD (*ln(Dist. To CBD)*) are significant and positive for the two models. Therefore, the greater the distance between the destination and CBD, the greater the propensity that an individual is likely to choose the destination to avoid congestion in the CBD. Finally, the McFadden's ρ^2 for the constrained model improved considerably from 0.08 in the unconstrained model to 0.467 in the constrained model. Therefore, the constrained model better fits the data compared to the unconstrained model.

Two other models for both the unconstrained and constrained shopping destination models are estimated using different specifications as summarized in Table 5-14. The specifications for the second unconstrained model are similar to the previous model, but more interaction terms between the destination attributes and the traveler's socio-economic and demographic characteristics are included. The results show that the destinations' attractiveness vary according to the individuals' characteristics (i.e., age, gender, employment, and household size). The inclusion of the interaction terms improves the McFadden's ρ^2 slightly.

As for the results of the 2nd constrained model, the coefficients of the store categories vary considerably compared to the unconstrained model with *Walmart* and *Costco* being the most influential store categories. Similar to the previous models, the coefficients of the *ln(Dist. To HWY)* and the *ln(Dist. To CBD)* variables decrease and increase the traveler's propensity to choose that destination, respectively. Also, the results show that female travelers are more sensitive to the *ln(Dist. To CBD)* compared to males. In addition, instead of including the traveler's *Maximum Travel Window* as a variable on its own, it is broken down into categories depending on the type of the visited destination (i.e., *Grocery Store*, *Department Store*, *Walmart*, *Costco*, *Shopping Mall*, *Shoppers Drug Mart*, and *Other Pharmacies*). The results show that although the coefficients of *Maximum Travel Window* based on the destination type are all positive and significant, they vary substantially.

The inclusion of the interaction terms in the model revealed how the attractiveness of the modeled destinations vary with respect to the traveler's characteristics. For instance, an individual who is 20-34 years old has a lower probability of choosing *Other Pharmacies* as a destination. By comparison, an individual who is 20-49 years old is more likely to select *Walmart* as a destination. McFadden's ρ^2 values suggest that the constrained model is still a better fit for the data.

Table 5-12 Results for the unconstrained destination choice models for the survey and 5% sample

Variable	Survey - Unconstrained Model			5% Sample - Unconstrained Model 1		
	Coeff.	Error	t-stat	Coeff.	Error	t-stat
Food Store	0.6	0.05	10.98	0.49	0.04	11.56
Department Store	0.36	0.06	6.49	0.26	0.04	6.03
Walmart	2.18	0.07	30.78	1.84	0.06	33.16
Costco	0.55	0.19	2.81	0.87	0.14	6.05
Shoppers Drug Mart	1.05	0.06	18.72	0.86	0.05	18.92
Other Pharmacies	-0.68	0.1	-6.64	-0.62	0.08	-7.38
ln (Dist. to Hwy)	-0.26	0.03	-9.5	-0.17	0.02	-7.79
ln (Dist. to CBD)	1.32	0.11	12.28	1.16	0.08	13.82
MTW	0.01	0.01	0.88	0.002	0.01	0.36
MTW × Walked	-0.04	0.02	-2.08	-0.04	0.02	-2.5
Age (<20) × Shoppers Drug Mart	-1.85	1.03	-1.81	-1.72	0.73	-2.37
Age (20-49) × Department Store	0.29	0.11	2.56	0.46	0.09	4.99
Employed × Department Store	-0.24	0.15	-1.56	-0.22	0.12	-1.86
Number of Observations	1,946			3,160		
Log Likelihood Function (β)	-8,613.18			-13,705.66		
Log Likelihood Function (0)	-9,426.79			-14,645.74		
ρ^2	0.086			0.064		

***MTW stands for Maximum Time Window**

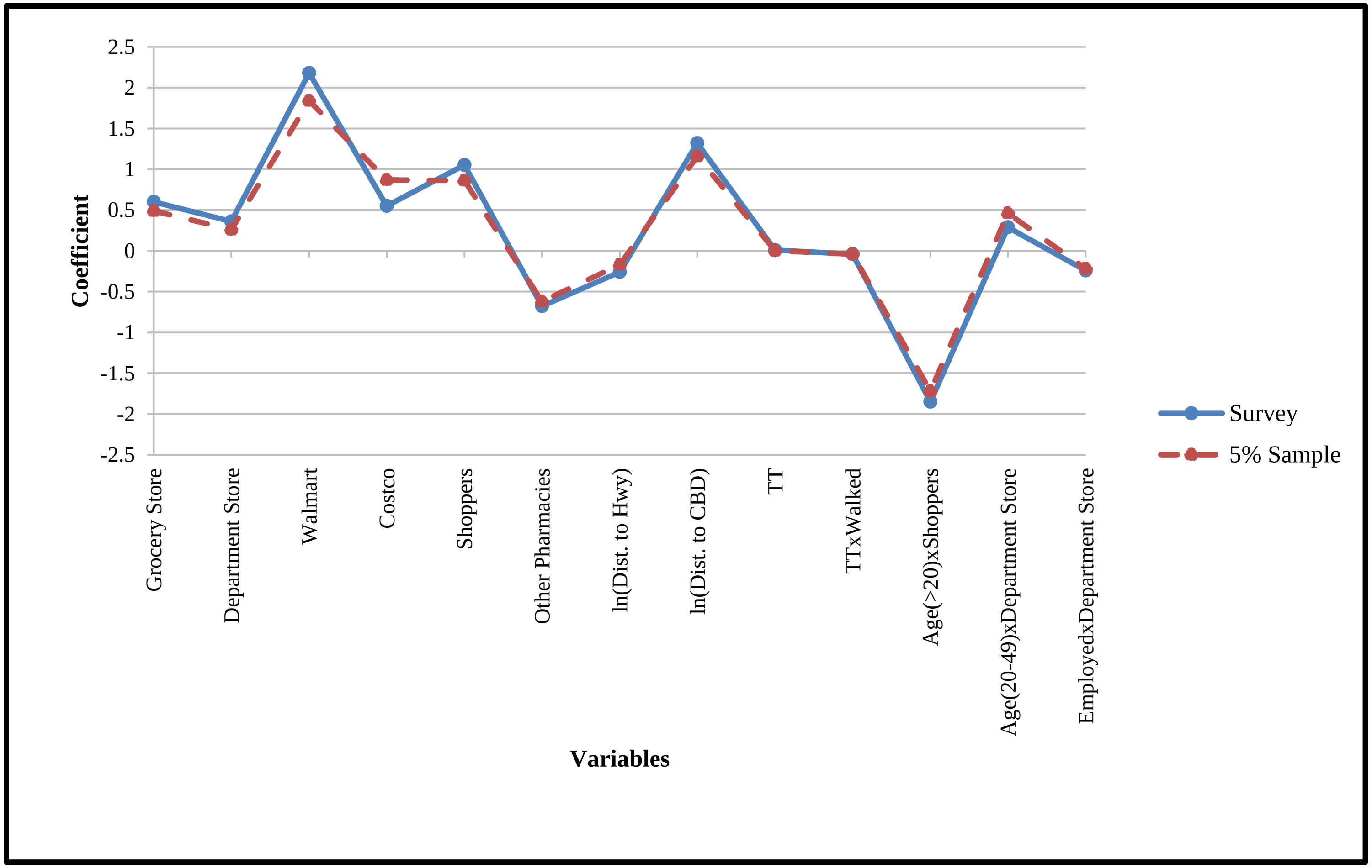

Figure 5-3 Comparison between the survey and 5% sample model coefficients

Table 5-13 Results for the unconstrained and constrained destination choice models 1

Variable	5% Sample - Unconstrained Model 1			5% Sample - Constrained Model 1		
	Coeff.	Error	t-stat	Coeff.	Error	t-stat
Grocery Store	0.61	0.04	13.72	0.65	0.05	13.01
Department Store	0.65	0.05	13.33	0.60	0.05	11.11
Walmart	1.61	0.06	28.03	1.68	0.07	24.44
Costco	1.34	0.15	9.17	1.25	0.17	7.58
Shopping Mall	1.35	0.06	22.81	1.36	0.07	19.26
Shoppers Drug Mart	0.63	0.05	13.48	0.73	0.05	13.42
Other Pharmacies	-0.39	0.08	-4.61	-0.39	0.09	-4.43
ln(Dist. to Hwy)	-0.13	0.02	-5.62	-0.10	0.03	-3.73
ln(Dist. to CBD)	1.01	0.08	11.92	0.97	0.1	10.09
MTW	0.004	0.01	0.70	2.14	0.05	46.33
MTW × Walked	-0.04	0.02	-2.44	0.17	0.15	1.15
Age(<20) × Shoppers Drug Mart	-1.72	0.73	-2.37	-0.08	0.15	-0.53
Age(20-49) × Department Store	0.46	0.09	5.00	0.47	0.11	4.45
Employed × Department Store	-0.22	0.12	-1.87	-0.14	0.14	-0.99
Number of Observations	3,160			3,160		
Log likelihood function (β)	-13,469.34			-6,445.87		
Log likelihood function (0)	-14,645.74			-12,084.42		
ρ^2	0.08			0.467		

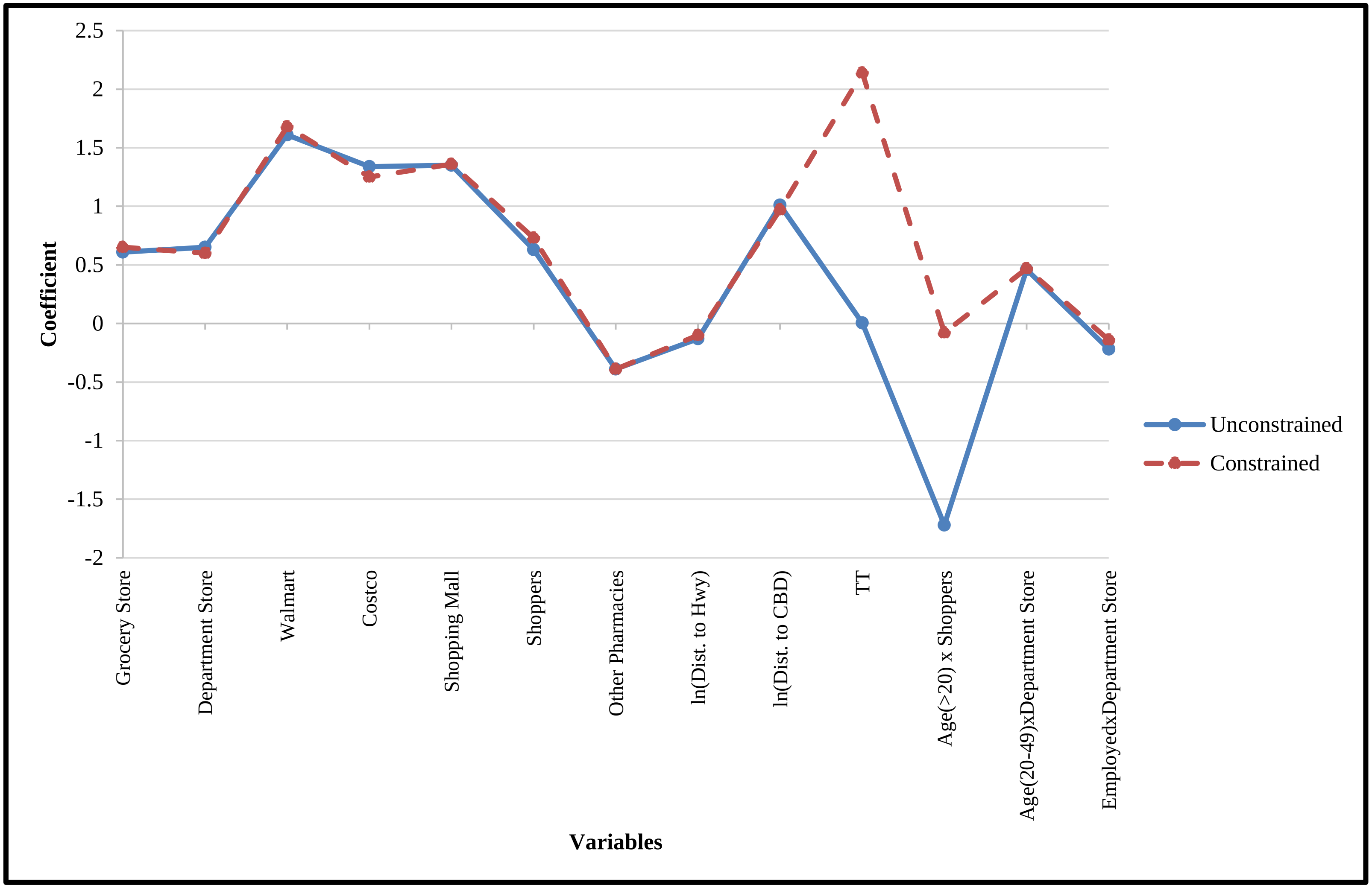

Figure 5-4 Comparison between the unconstrained and constrained 5% sample model coefficients

Table 5-14 Results for the unconstrained and constrained destination choice models 2

Variable	5% Sample - Unconstrained Model 2			5% Sample - Constrained Model 2		
	Coeff.	Error	t-stat	Coeff.	Error	t-stat
Grocery Store	0.61	0.04	13.77	-3.90	0.23	-16.96
Department Store	0.52	0.08	6.75	-3.08	0.17	-17.72
Walmart	1.54	0.13	12.22	1.33	0.17	7.71
Costco	1.53	0.15	9.96	1.35	0.23	5.82
Shopping Mall	1.24	0.09	14.17	-0.70	0.24	-2.89
Shoppers Drug Mart	0.63	0.05	13.55	-0.72	0.25	-2.86
Other Pharmacies	-0.39	0.08	-4.54	-4.95	0.32	-15.48
ln(Dist. to Hwy)	-0.12	0.02	-5.45	-0.26	0.03	-8.13
ln(Dist. to CBD)	1.01	0.08	11.96	1.32	0.17	7.59
Female × ln(Dist. to CBD)	--	--	--	0.44	0.22	1.97
MTW	0.01	0.01	0.57	--	--	--
MTW × Walked	-0.04	0.02	-2.45	--	--	--
MTW × Grocery Store	--	--	--	0.72	0.02	30.37
MTW × Department Store	--	--	--	0.50	0.02	29.54
MTW × Shopping Mall	--	--	--	0.27	0.03	10.12
MTW × Shoppers Drug Mart	--	--	--	0.05	0.02	2.06
MTW × Other Pharmacies	--	--	--	0.60	0.03	18.49
Age(<20) × Shoppers Drug Mart	-1.66	0.73	-2.28	--	--	--
Age (20-34) × Other Pharmacies	--	--	--	-2.54	0.59	-4.34
Age(20-49) × Department Store	0.33	0.1	3.3	--	--	--
Age(20-49) × Walmart	--	--	--	0.61	0.21	2.95
Age(35-49) × Grocery Store	--	--	--	-0.68	0.28	-2.43
Age(35-49) × Shoppers Drug Mart	--	--	--	0.78	0.29	2.7
Age(50-64) × Costco	-1.15	0.43	-2.64	-3.38	0.67	-5.04
Age(65+) × Department Store	--	--	--	-0.16	0.1	-1.51
Age(65+) × Walmart	-0.29	0.1	-2.88	--	--	--
Female × Shopping Mall	0.18	0.1	1.86	--	--	--
Employed × Grocery Store	--	--	--	1.14	0.28	3.99
Employed × Department Store	1.7	0.49	3.49	--	--	--
Household Size × Grocery Store	--	--	--	-0.48	0.07	-6.90
Household Size × Department Store	--	--	--	0.09	0.04	2.42
Household Size × Walmart	0.06	0.03	1.89	0.24	0.07	3.54
Household Size × Costco	0.08	0.03	2.38	--	--	--
Number of Observations	3,160			3,160		
Log likelihood function (β)	-13,445.50			-4,598.21		
Log likelihood function (0)	-14,645.74			-8,491.21		
ρ^2	0.082			0.458		

5.2.1.3 Validation of Destination Choice Models

A comparison of the performance between the unconstrained and constrained models is conducted based on out-of-sample predictions. As stated in the method of analysis chapter, the two measures considered are the percent predicted right and the margin of error in minutes between the predicted and actual destinations. Table 5-15 summarizes the results of the first measure. It is clear that the constrained model 2 has the highest percent predicted right with a value of 68.7%. On the other hand, the unconstrained models show very poor predictive ability with the highest percent predicted right being 6.4%. Therefore, the 5% Sample-Constrained Model 2 is the best destination choice model among all the estimated models.

Table 5-15 Out-of-sample validation results

Models	% Predicted Right
Survey - Unconstrained Model	3.80%
5% Sample - Unconstrained Model 1	6.40%
5% Sample - Unconstrained Model 2	5.70%
5% Sample - Constrained Model 1	30.30%
5% Sample - Constrained Model 2	68.70%

The results pertaining to the second validation measure (i.e., margin of error) are presented in Table 5-16 and Figure 5-5 for the unconstrained models, and Table 5-17 and Figure 5-6 for the constrained models. The margin of error suggests that the difference between the actual and predicted destinations is not minor in the case of the unconstrained models 1 and 2. More than 85% of the travelers' predicted destinations have a margin of error of five minutes or greater. On the other hand, for the constrained model 2, more than 81% of the traveler's predicted destinations have a margin of error of less than five minutes and about 69% of the predicted destinations have a margin of error of less than one minute. In conclusion, the validation results based on the margin of error metric also favours the *5% Sample - Constrained Model 2* over the other models.

Table 5-16 Margin of error results in minutes for the unconstrained models

Margin of Error (min)	5% Sample - Unconstrained Model 1	5% Sample - Unconstrained Model 2
0	84	84
1	93	98
2	10	10
3	121	121
4	124	115
5	469	467
6	76	78
7	433	426
8	2	17
9	276	273
10	112	105
11	391	382
12	451	455
13	127	117
14	220	241
18	2	2
Total	2,991	2,991

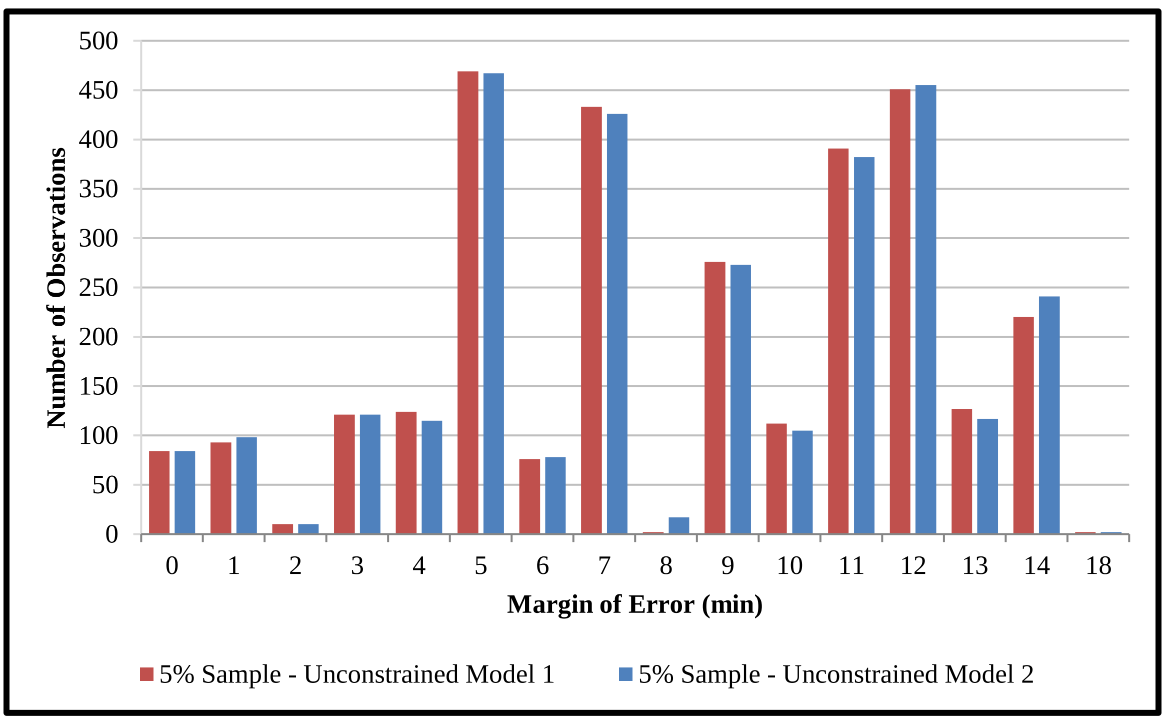

Figure 5-5 Margin of error distribution for the unconstrained models

Table 5-17 Margin of error results in minutes for the constrained models

Margin of Error (min)	5% Sample - Constrained Model 1	5% Sample - Constrained Model 2
0	1,047	2,055
1	201	40
2	140	82
3	118	96
4	156	164
5	315	109
6	203	120
7	161	96
8	205	74
9	79	75
10	70	80
11	59	0
12	137	0
13	44	0
14	40	0
15	10	0
16	3	0
18	3	0
Total	2,991	2,991

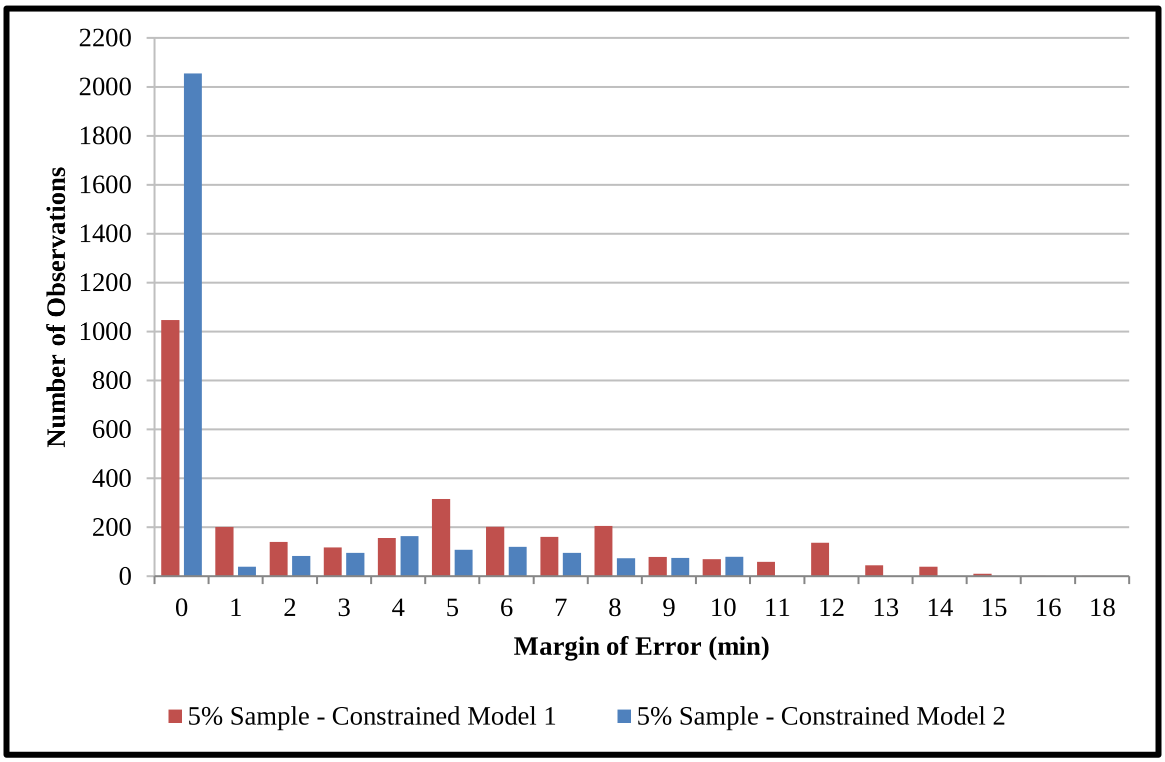

Figure 5- 6 Margin of error distribution for the constrained models

5.2.2 Micro Analysis – Location of TAZ as Destinations

Using the unconstrained and constrained TAZ choice sets, two MNL models are estimated to analyze and predict the destination TAZ chosen by each traveler. The estimation results are shown in Table 5-18. Similar to the micro-based destination choice models, the coefficient of each store category reveals the contribution of each specific store type to the destination utility. Compared to the micro destination models, more store categories are significant in these models. In addition, most of the store categories increase the TAZs' attractiveness. In the case of the unconstrained model, the large magnitudes of *Shoe Stores* and *Department Stores* demonstrate their influential roles in the utility functions. On the other hand, the store categories that are most positively influential in the constrained model are *Walmart* and *Shoppers Drug Mart*.

The coefficients for the natural logarithm of the destination's distance to the CBD [*ln(Dist. To CBD)*] have opposite effects in the unconstrained and constrained models. In the unconstrained model, the *ln(Dist. To CBD)* is significant and positive, while its significant and negative in the constrained model. This is most likely because the *Maximum Travel Window* is not included in the estimation of the unconstrained model. The exclusion of the *Maximum Travel Window* variable was necessary since it causes the model to misbehave. As expected, the *Maximum Travel Window* coefficient in the constrained model is positive and highly significant. The models also include interaction terms to reveal the variation of the TAZs' attractiveness with respect to the traveler's characteristics. According to the achieved McFadden's ρ^2 values, the constrained model is by far a better fit for the data. In addition, the constrained model's predictive ability is superior over the unconstrained model. The validation indicates that the percentage predicted right in the constrained model is 81%. This is compared to only 12% predicted right in the unconstrained model.

Table 5- 18 Results for the unconstrained and constrained TAZ destination choice models

Variable	Unconstrained Model			Constrained Model		
	Coeff.	Error	t-stat	Coeff.	Error	t-stat
Grocery Store	0.29	0.06	4.88	0.66	0.10	6.50
Walmart	0.88	0.08	11.37	1.86	0.15	12.76
Shoppers Drug Mart	0.73	0.05	14.52	1.55	0.09	16.95
Other Pharmacies	0.32	0.07	4.31	0.19	0.13	1.49
Department Store	1.01	0.07	13.56	0.40	0.12	3.37
Variety Store	0.05	0.05	1.05	-0.29	0.08	-3.54
Hardware Store	-0.04	0.06	-0.66	-0.25	0.09	-2.75
Auto and Home Supply Store	0.04	0.06	0.62	-0.38	0.09	-3.99
Women Clothing Store	-0.59	0.08	-7.30	-1.33	0.16	-8.35
Shoe Store	1.43	0.10	14.31	0.89	0.16	5.62
Furniture Store	-1.00	0.13	-7.92	-1.03	0.18	-5.58
Shopping Mall	0.20	0.06	3.56	0.58	0.11	5.36
Alcoholic Beverages	-0.03	0.05	-0.65	-0.40	0.08	-4.75
Commercial Banks	0.43	0.07	6.54	0.22	0.10	2.06
ln(Distance to CBD)	2.13	0.10	20.46	-0.46	0.18	-2.52
Female × ln(Distance to CBD)	0.29	0.13	2.21	0.36	0.21	1.68
MTW	--	--	--	2.92	0.11	25.50
MTW × Female	--	--	--	0.63	0.15	4.21
Age(20-49) × Shoppers Drug Mart	-0.18	0.09	-1.95	-0.37	0.17	-2.19
Age(20-34) × Department Store	-0.23	0.14	-1.63	0.38	0.24	1.56
Age(20-34) × Auto and Home Supply Store	-0.28	0.15	-1.89	-0.79	0.27	-2.90
Age(35-49) × Walmart	0.24	0.16	1.50	1.12	0.30	3.75
Age(35-49) × Hardware Store	-0.30	0.13	-2.34	-0.69	0.21	-3.24
Age(35-49) × Women Clothing Store	0.03	0.18	0.17	-1.09	0.39	-2.81
Age(35-49) × Shopping Mall	-0.27	0.14	-1.87	0.63	0.28	2.28
Age(35-49) × Alcoholic Beverages	0.25	0.13	1.98	0.73	0.22	3.31
Age(65+) × Other Pharmacies	0.17	0.08	2.20	0.39	0.14	2.82
Age(65+) × Department Store	-0.23	0.08	-2.72	-0.27	0.14	-1.98
Number of Observations	3,160			3,160		
Log likelihood function (β)	-10245.34			-7460.23		
Log likelihood function (0)	-12233.00			-1968.13		
ρ^2	0.16			0.74		
% Predicted Right	12.20%			81.40%		

CHAPTER 6: CONCLUSIONS

This thesis advances the traditional urban transportation modeling system (UTMS) by adopting the micro-based paradigm to study travel demand behavior in the London Census Metropolitan Area (CMA), Ontario. It does so by focusing on developing improved models for the trip generation and trip distribution modules of the UTMS. Two data sources were employed in the analysis: 1) the 2011 Canadian Census, and 2) a conventional Household Travel Survey conducted for London in 2009. Using the latter data, a complete list of over 195,000 households is synthesized using the combinatorial optimization population synthesis technique.

The first objective of the analysis is to compare various techniques that could be used to model trip generation (i.e., regression, cross-classification, discrete choice, and count models) at the micro-level. Also, compare the predictive ability of these micro-models against conventional zone-based models. The second objective is to apply advanced geo-spatial methods and statistical techniques to model trip distribution using micro-data from the London Household Travel Survey (LHTS). To date, trip distribution in the four-stage model has relied on the gravity approach, which is too simplistic for capturing the complexities of spatial interaction within a travel demand model.

6.1 Trip Generation

Trip generation is a critical step in the urban transportation modeling system. In this study, work and non-work trip frequency data for the London CMA, Ontario are analyzed with the use of advanced micro-based techniques. For each trip purpose, four micro-based trip generation models (linear-regression, cross-classification, ordered choice, and count models) are estimated and compared. In addition, the predictive ability of these micro-models is compared against a conventional zone-based model for work trips. Several socio-demographic factors including household size, age, gender, and vehicle ownership are used to estimate the models. Also, dummy variables for the hour of day when the trip took place are introduced.

The results of the models show that work and non-work trip frequency differs by trip purpose, hour of day, and household characteristics. First, work trip frequency increases as age increases, reaching a peak for the working age group, and then decreases for the senior age group. In addition, males, vehicle ownership, and AM-peak period are found to have a positive impact on work trip frequency. The estimates from the linear-regression, ordered logit and Poisson models are consistent in terms of behavior and predictive ability. Surprisingly, the cross-classification has the weakest results.

In general, non-work trip frequency increases as age increases, reaching a peak for the senior age group. The hour of the day dummy variables were positive and highly significant for the periods between 9 am to 12 pm, and 1 pm to 2 pm. Furthermore, using the Social Dummy in the model specification proves to be significant. Compared to shopping trips, social trips are more likely to increase the non-work trip frequency per household. In terms of gender, females are more likely to generate social trips compared to males. When considering the models' predictive ability, the estimations from the linear-regression, ordered logit and Poisson models are consistent as in the case of work trips. However, the cross-classification model resulted in inferior predictions. Trip rates vary considerably for the different household structures and vehicle ownership, but no clear pattern is observed. Hence, like the work-trips case, the cross-classification has the weakest results. In summary, caution should be practiced when using cross-classification to model trip generation. While the result among the three statistical techniques were consistent, the use of linear regression models is not advisable if the models are to be used in a prediction exercise. This is because a linear regression model runs the risk of producing negative predictions. On the other hand, while a Poisson regression model is more suitable than a linear regression model (the generated trip per household could be treated as a count), we believe an ordered logit model is even more suitable for predictions. As shown in Table 4-3 in Chapter 4, 35% and 30% of the modeled households generated zero work and non-work trips, respectively. Therefore, a Poisson regression model should probably be expanded to a Zero-Inflated Count model to account for the large percentage of zero counts in the dependent variable. Also, the use of a Poisson regression runs the risk of producing very large counts (i.e., trip frequencies) for certain households. Given the ordered nature of the generated trips per household (as shown in Table 4-3), the ordered logit model appears to be the most appropriate. The model also overcome the inherited limitations of the linear-regression and Poisson regression models when performing predictions.

6.2 Trip Distribution

Trip distribution choice models are estimated for shopping trips at the micro-level in two different ways. The first way considers the actual locations of the stores as the possible destinations in the choice sets. The second way considers the locations of the TAZs where the trips ended as the possible destinations. For the first set of models, a MNL model is first estimated to model and predict the individual's type of visited destination (i.e., food, medicine, or other). Using the individual's maximum travel window (i.e., calculated travel time) and destination type, constrained choice sets are formulated for each individual separately. Next, unconstrained and constrained MNL models are estimated to test the influence of various factors (namely: store type categories, destination attractiveness categories, and socio-economic and demographic characteristics) on the unconstrained and constrained destination choice sets.

The use of the various store categories in the unconstrained final model indicates that most of them increase the destinations' utility functions, with Walmart and Shopping Malls have the larger magnitudes, demonstrating their influential roles as a shopping destination. Coefficient of each store category reveals the contribution of each specific store type to the destination utility. As for the destination attractiveness categories, the natural logarithm of the destinations distance to a highway is found to have a negative effect on the destinations' utility. While, the natural logarithm of the destinations distance to the CBD has a positive influence on the destination's utility. On the other hand, travel time is insignificant in the model. Finally, the socio-economic and demographic characteristics are introduced as interaction terms with the store types and the destination attributes. Therefore, revealing the variations in behavior depending on the individual's characteristics (i.e., age, gender, employment, and household size).

For the constrained final model, the store categories effects vary considerably with Walmart and Costco being the most influential. For the destination attractiveness categories, the same effects are seen as in the unconstrained model. However, in this model instead of introducing the *Maximum Travel Window* as a variable on its own, it is interacted with the store categories to capture the variation of the effect of travel time on each store category separately. This proves beneficial as the *Maximum Travel Window* coefficients with respect to the store categories vary significantly. In the same fashion as the other models, the interaction terms are used to reveal how the destinations' attractiveness and store types vary with respect to the individual's characteristics. McFadden's Rho-squared and the validation results prove that the constrained model is superior to the unconstrained model. Hence, these results support our earlier statement on the importance of restricting the individual's choice set to reduce the bias in the models.

Considering the second set of trip distribution models, two MNL models are estimated, unconstrained and constrained choice sets where the TAZs are the destinations. The variables considered in these models are the same as in the previous models. The results reveal that more instore categories are significant and increase the destinations' attractiveness compared to the previous models. The unconstrained models show that Shoe Stores and Department Stores are the most positively influential compared to the other store types. On the other hand, the constrained models illustrate that Walmart and Shoppers Drug Mart are the most positively influential.

The destination attractiveness categories show that the natural logarithm of the destination's distance to the CBD has positive and negative influence in the unconstrained and constrained models, respectively. In addition, the range of service area is only considered in the constrained trip generation model. The range of service area is seen to increase the destinations' attractiveness due to the method considered in constraining the choice sets. The results also reveal the variation of the destinations' attractiveness with respect to the traveler's characteristics. The models' goodness-of-fit and the predictive ability prove that the constrained model is favored with an 81% predicted right compared to 12% predicted right for the unconstrained model. When comparing the two sets of constrained models, the model that considers the TAZs as the destinations is preferred as it is capable of providing better predictions.

6.3 Contributions and Policy Implications

The analysis presented here offers a pioneering effort to address an important gap in current transportation research in terms of trip generation and trip distribution modeling. The contributions of this thesis are as follows: 1) it advances the current state of knowledge on travel demand modeling; 2) it compares the various techniques that could be used to model trip generation for work and non-work trips; and 3) it applies advance geo-spatial methods and statistical techniques to model shopping trip distribution, something that has not been investigated adequately in literature. Another very important aspect in this study is the comparison between the micro- and macro-models in the case of trip generation.

From a transport policy perspective, modeling and understanding travel demand is essential for both urban transport and land-use planning, since trip generation and trip distribution influence traffic and level-of-service on the transportation network. The results from this research can benefit planners and decision-makers, as it will allow for the prediction of passenger vehicle movements in London CMA in future years at the micro-level. This will provide the opportunity to accurately determine the future demand on the transportation network. Refined predictions will ensure that adequate transportation facilities are available to meet future travel demand. Such predictions will also help local governments to plan for infrastructure maintenance to ensure high-level of service. Sustainable transportation policies can then be created to improve passenger vehicle movements in the CMA. In addition, the approach devised in this thesis can help overcome and eliminate bias in existing trip distribution models by constraining the individuals' choice sets according to accessibility measures (i.e., range of service area), destination characteristics, and/or socio-economic and demographic characteristics. In a nutshell, the research conducted in this thesis highlights some of the drawbacks in existing tradition urban transportation modeling systems and offers a platform for performing better predictions using data derived from conventional household travel surveys.

6.4 Limitations and Recommendations

Finally, the limitations of this work relate to the 2009 London Household Travel Survey. Although the survey offers valuable data in terms of households and individuals characteristics in the study area, little information is included on the actual activity type and activity duration for the different trip purposes. Therefore, all the reported destinations are explored and analyzed separately, which is very time consuming and not always accurate. Also, most of the records in the survey did not reveal information on their household income. Hence, the models did not test the influence of income on trip generation and trip distribution. In addition, the shopping trips from the survey mainly concentrate on the City of London compared to the sub-urban areas. The latter does not allow this study to compare the urban and suburban behavior. Nevertheless, these limitations can be overcome in future household travel surveys.

Future research on this topic should first complete the last two stages of the four-stage model; mode choice (Khan et al., 2014; Kiamura, 2009) and traffic assignment. By devising an improved four-stage model that makes use of a conventional Household Travel Survey for London, Ontario exclusive values of accessibility can be derived. Second, future research should focus on estimating trip distribution models for other trip purposes not only shopping trips. Also, advanced discrete choice modeling techniques such as the Mixed Logit Model (MXL) could be utilized to estimate the destination choice model. It is also important to note that when it comes to the implementation of the estimated trip distribution models to perform predictions, the same framework can be applied. However, an extra model to predict the maximum time window for each traveler will be required. Such model could benefit from the work conducted by Maoh and Tang (2012).

REFERENCES

Arliansyah, J. and Hartono, Y. (2015) 'Trip attraction model using radial basis function neural networks', *Procedia Engineering*, 125, pp. 445–451.

Axhausen, K. and Garling, T. (1992) 'Activity-based approaches to travel analysis: Conceptual frameworks, models, and research problems'. *Transportation Reviews*, 12(4), 323-341.

Badoe, D.A. (2007) 'Forecasting travel demand with alternatively structured models of trip frequency', *Transportation Planning and Technology*, 30(5), pp. 455–475.

Badoe, D.A. and Chen, C.C. (2004) 'Unit of analysis in conventional trip generation modelling: An investigation', *Canadian Journal of Civil Engineering*, 31(2), pp. 272–280.

Ben-Akiva, M.E. and Lerman, S. (1985) 'Discrete Choice Analysis: Theory and Application to Travel Demand', *MIT Press, Cambridge, MA.*

Bhat, C.R. and Lawton, K.T. (2000) 'Passenger Travel Demand Forecasting', *Transportation Research Board, National Research Council (USA), Millennium Paper.*

Canada, E.G. of and Canada, C.C. (2016) 'Environment and climate change Canada - environmental indicators - greenhouse gas emissions by economic sector'. [Online]. Available at:

https://ec.gc.ca/indicateurs-Indicators/default.asp?lang=en&n=F60DB708-1 [Accessed: 20 October 2016].

Castiglione, J., Bradley, M. and Gliebe, J. (2015) ‘Activity-Based Travel Demand Models’, [Online]. Available at: http://onlinepubs.trb.org/onlinepubs/shrp2/SHRP2_C46.pdf [Accessed: 20 October 2016].

Chang, J.S., Jung, D., Kim, J. and Kang, T. (2014) ‘Comparative analysis of trip generation models: Results using home-based work trips in the Seoul metropolitan area’, *Transportation Letters*,6(2) , pp. 78–88.

Clifton, K., Muhs, C., Morrissey, S., Morrissey, T., Currans, K. and Ritter, C. (2013) ‘Examining consumer behavior and travel choices’. [Online]. Available at: http://kellyjclifton.com/Research/EconImpactsofBicycling/OTRECReport-ConsBehavTravelChoices_Nov2012.pdf [Accessed: 20 October 2016].

Cubukcu, K.M. (2001) ‘Factors Affecting Shopping Trip Generation Rates in Metropolitan Areas’, *Studies in Regional and Urban Planning*

DMTI Spatial (2016). [Online]. Available at: http://www.dmtispatial.com/canmap/ [Accessed: 18 April 2017].

Ettema, D.F. and Timmermans H. (1997) ‘Activity-based approaches to travel analysis’, *Pergamon/Elsevier, Amsterdam*.

Fox, J. and Patruni, B. (2015) ‘STM3 2011 base frequency, mode-destination and car ownership models’. Available at:

http://www.rand.org/pubs/research_reports/RR1125.html [Accessed: 17 October 2016].

Greene, W., and Hensher, D. (2003) 'A latent class model for discrete choice analysis: Contrasts with mixed logit', *Transportation Research Part B: Methodological*, pp. 681-698.

Griesenbeck, B. and Garry, G. (2007) 'Comparison of activity-based tour model to Four-Step model as a tool for metropolitan transportation planning'. [Online]. Available at: http://trbappcon.org/2007conf/papers/session08/05-Comparison_of_ActivityBased_Tour_Model_to_FourStep_Model.pdf [Accessed: 20 October 2016].

Hanson, S. (1980) 'The Importance of the Multi-Purpose Journey to Work in Urban Travel Behavior', *Transportation*, 9, pp. 229–248.

Hanna, M. and Tang, Z. (2012) 'Determinants of normal and extreme commute distance in a sprawled midsize Canadian city: evidence from Windsor, Canada', *Journal of Transport Geography*, 25, pp. 50-57.

Horner, M.W. and O'Kelly, M.E. (2007) 'Is non-work travel excessive?', *Journal of Transport Geography*, 15(6), pp. 411–416.

Huntsinger, L.F., Rouphail, N.M. and Bloomfield, P. (2013) 'Trip generation models using cumulative logistic regression', *Journal of Urban Planning and Development*, 139(3), pp. 176–184.

Jang, T.Y. (2005) ‘Count data models for trip generation’, *Journal of Transportation Engineering*, 131(6), pp. 444–450.

Khan, S., Maoh, H., Lee, C. and Anderson, W. (2014) ‘Toward sustainable urban mobility: Investigating nonwork travel behavior in a sprawled Canadian city’, *International Journal of Sustainable Transportation*, 10(4), pp. 321–331.

Kim, J. and Lee, S. (2017) ‘Comparative analysis of traveler destination choice models by method of sampling alternatives’, *Transportation Planning and Technology*, 40(4), pp.465 –478.

Kiamura, R. (2009) ‘A dynamic model system of household car ownership, trip generation, and modal split: model development and simulation experiment’ *Transportation*, 36, pp. 711–732.

Kitamura, R., Chen, C., Pendyala, R.M., and Narayanan, R. (2000) ‘Micro-simulation of daily activity-travel patterns for travel demand forecasting’, *Transportation*, 27, pp. 25–51.

Krizek, K.J. (2003) ‘Neighborhood services, trip purpose, and tour-based travel’, *Transportation*, 30(4), pp. 387–410.

Goulias, K.G., Pendyala, R.M., and Kitamura, R. (1990) ‘Practical Method for The Estimation of Trip Generation and Trip Chaining’, *Transportation Center, University of California.*

Greene, W.H. (2011) ‘LIMDEP Version 9 Student Reference Guide’. [Online]. Available at: http://pages.stern.nyu.edu/~wgreene/DiscreteChoice/Software/LIMDEP-Short-Student-Manual.pdf [Accessed: 18 April 2017].

Longden, T. (2016) ‘The regularity and irregularity of travel: An analysis of the consistency of travel times associated with subsistence, maintenance and discretionary activities by Thomas Longden: SSRN’.

Ma, J., and Goulias, K.G. (1999) ‘Application of Poisson Regression Models to Activity Frequency Analysis and Prediction’, *Transportation Research Record*, 1676, pp. 86–94.

Malayath, M. and Verma, A. (2013) ‘Activity based travel demand models as a tool for evaluating sustainable transportation policies’, *Research in Transportation Economics*, 38(1), pp. 45–66.

Mazzulla, G., and Forciniti, C. (2012) ‘Spatial association techniques for analyzing trip distribution in an urban area’, *Eur. Transp. Res. Rev.*, 4, pp. 217–233.

McFadden, D. (1978). ‘MODELING THE CHOICE OF RESIDENTIAL LOCATION’, *Transportation Research Record*. [Online]. (673). Available at: https://trid.trb.org/view.aspx?id=87722. [Accessed: 20 April 2016].

McNally, M.G. (2007) ‘The Four Step Model’, *Institute of Transportation Studies, University of California.*

Naess, P. (2012) ‘Urban form and travel behavior: Experience from a Nordic context’, *Journal of Transport and Land Use*, 5(2).

Neutens, T., Delafontaine, M., Scott, D.M. and De Maeyer, P. (2012) ‘An analysis of day-to-day variations in individual space–time accessibility’, *Journal of Transport Geography*, 23, pp. 81–91.

Omer, M., Kim, H., Sasaki, K., and Nishii, K. (2010) ‘A Tour-based Travel Demand Model Using Person Trip Data and Its Application to Advanced Policies’, *KSCE Journal of Civil Engineering*, 14(2), pp. 221–230.

Pozsgay, M. and Bhat, C. (2000) ‘Modeling Attraction-End Choice for Urban Recreational Trips: Implications for Transportation, Air Quality and Land-Use Planning’. [Online]. Available at: http://www.ce.utexas.edu/prof/bhat/reports/utc_report167800.pdf [Accessed: 24 October 2016].

Páez, A., Scott, D., Potoglou, D., Kanaroglou, P. and Newbold, K.B. (2007) ‘Elderly mobility: Demographic and spatial analysis of trip making in the Hamilton CMA, Canada’, *Urban Studies*, 44(1), pp. 123–146.

Roorda, M.J., Páez, A., Morency, C., Mercado, R. and Farber, S. (2010) ‘Trip generation of vulnerable populations in three Canadian cities: A spatial ordered probit approach’, *Transportation*, 37(3), pp. 525–548.

Ryan, J., Maoh, H. and Kanaroglou, P. (2009). ‘Population synthesis: Comparing the major techniques using a small, complete population of firms’, *Geographical Analysis*, 41 (2), pp. 181–203.

Saw, K., Katti, B.K. and Joshi, G.K. (2015) ‘Ab Temporal Trip Generation Modelling for Primary Activities: A Case Study Of Fast Growing Metropolitan City’, *IJTTE,* 5(2), pp. 120–133.

Scott, D.M. and He, S.Y. (2012) ‘Modeling constrained destination choice for shopping: A GIS-based, time-geographic approach’, *Journal of Transport Geography*, 23, pp. 60–71.

Shay, E. and Khattak, A.J. (2012) ‘Household travel decision chains: Residential environment, automobile ownership, trips and mode choice’, *International Journal of Sustainable Transportation*, 6(2), pp. 88–110.

Shobeirinejad, M., Veitch, T., Smart, J.C.R., Sipe, N. and Burke, M. (2013) ‘Destination choice decisions of retail travelers: results from discrete choice modelling in Brisbane’, *Australasian Transport Research Forum 2013 Proceeding.*

Statistics Canada (2011). *2011 National Household Survey*. [Online]. Available at: http://dc.chass.utoronto.ca/cgibin/census/2011nhs/displayCensus.cgi?year=2011&geo=ct. [Accessed: 15 September 2016].

Su, F., Schmöcker, J.-D. and Bell, M.G.H. (2009) ‘Mode choice of older people before and after shopping: A study with London data’, *Journal of Transport and Land Use,* 2(1).

Tang, Zhongyuan, (2011) ‘Modeling Commuting Behaviour in Canadian Metropolitan Areas: Investigation of Factors and Methods with Micro Data’, *Master’s thesis, University of Windsor*.

Timmermans, H.J.P. (1996) ‘A stated choice model of sequential mode and destination choice behaviour for shopping trips’, *Environment and Planning A*, 28, pp. 173–184.

Train, K.E. (2009) ‘Discrete Choice Methods with Simulation’, *Cambridge University Press*.

Wang, J. (2011) ‘A prism- and gap-based model of shopping destination choice’, *Master’s thesis, University of Toronto*.

Whalen, K.E., Páez, A. and Carrasco, J.A. (2013) ‘Mode choice of university students commuting to school and the role of active travel’, *Journal of Transport Geography*, 31, pp. 132–142.

Williamson, P., Birkin, M. and Rees, P.H. (1998) ‘The estimation of population microdata by using data from small area statistics and samples of anonymized records’, *Environment and Planning A,* 30(5), pp. 785–816.

Winkelmann, R. (2015) ‘Counting on count data models’, *IZA World of Labor*. [Online]. Available at: http://wol.iza.org/articles/counting-on-count-data-models [Accessed: 7 June 2016]

Zhang, L., Nasri, A., Hong, J.H. and Shen, Q. (2012) ‘How built environment affects travel behavior: A comparative analysis of the connections between land use and vehicle miles traveled in US cities’, *Journal of Transport and Land Use,* 5(3).

Zhong, M., Shan, R., Du, D. and Lu, C. (2015) 'A comparative analysis of traditional four-step and activity-based travel demand modeling: A case study of Tampa, Florida', *Transportation Planning and Technology*, 38(5), pp. 517–533.

APPENDICES

Appendix A: Survey Data Description

Table A-1 Household size distribution

Household Size	Number of Households
1	1,082
2	2,296
3	1,281
4+	1,609
Grand Total	6,268

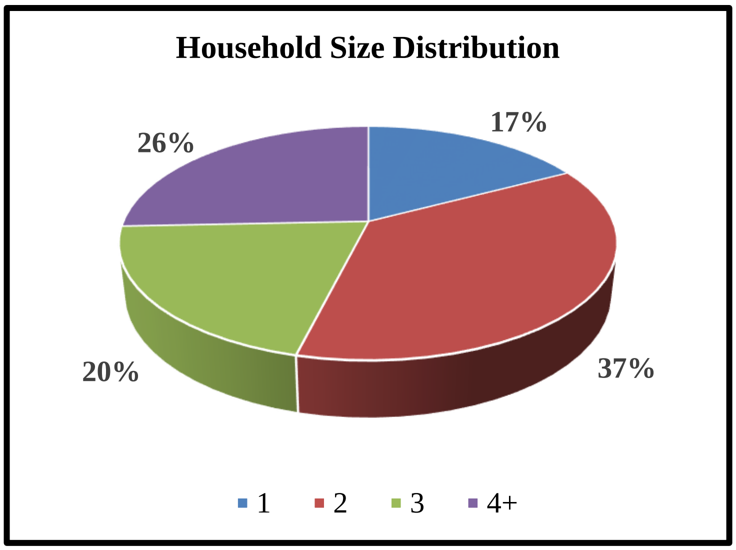

Figure A-1 Household size distribution

Table A- 2 Vehicle ownership distribution

Vehicle Ownership	Number of Households
0	522
1	2,359
2	2,886
3+	501
Grand Total	6,268

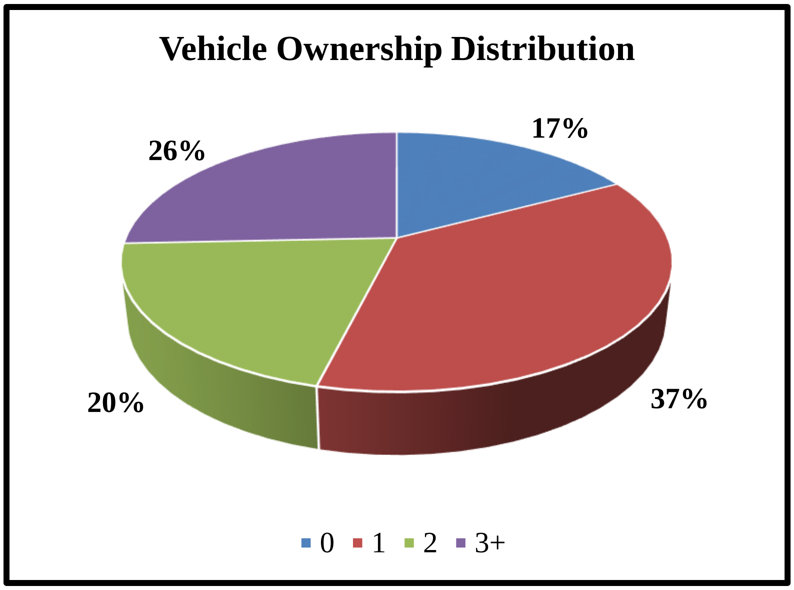

Figure A-2 Vehicle ownership distribution

Table A-3 Gender distribution in dataset

Gender	Number of Participants
Male	7,249
Female	7,227
Grand Total	14,476

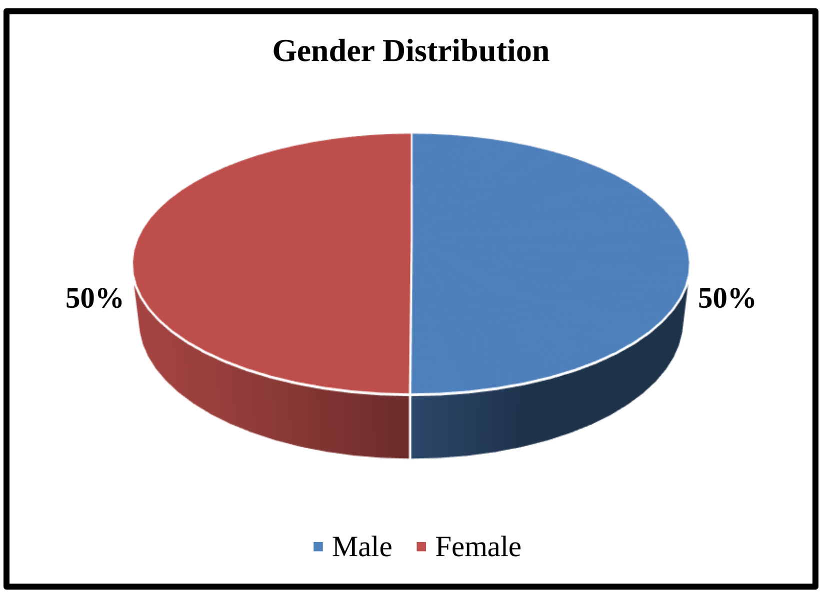

Figure A-3 Gender distribution in dataset

Table A-4 Age category distribution in dataset

Age Category	Number of Participants
Age1	2,125
Age2	3,360
Age3	4,268
Age4	2,640
Age5	2,083
Grand Total	14,476

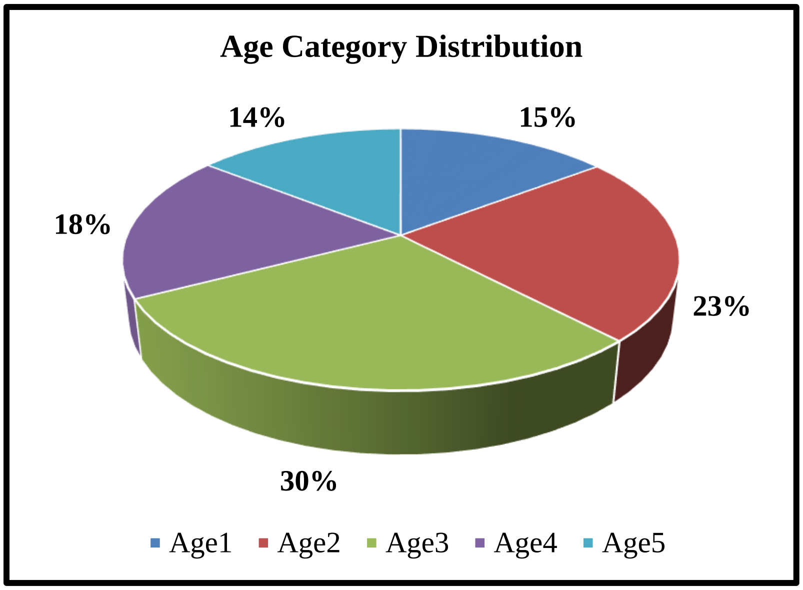

Figure A-4 Age category distribution in dataset

Table A- 5 Employment status distribution in dataset

Employment Status	Number of Participants
Employed-Full time	7,361
Employed-Part time	644
Not Employed	3,583
High School Students	1,619
Post Secondary Student	1,269
Grand Total	14,476

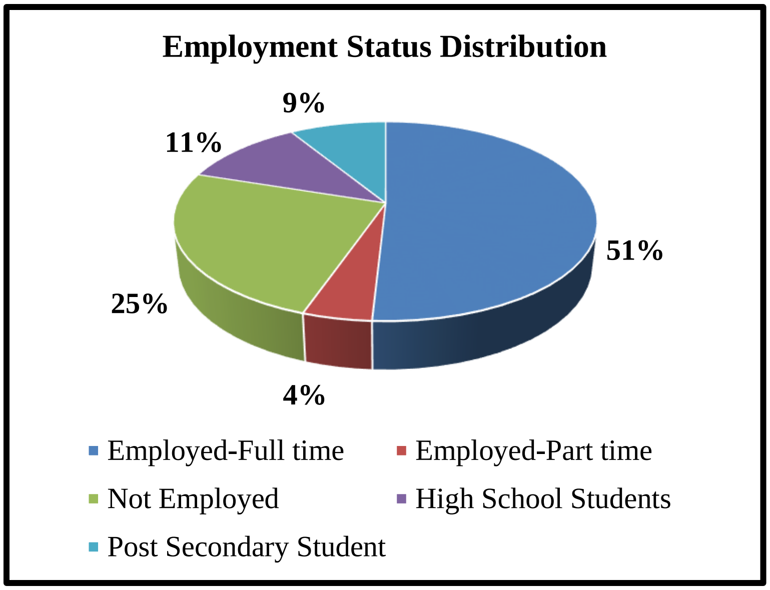

Figure A- 5 Employment status distribution in dataset

Table A- 6 Trip purpose distribution in dataset

Trip Purpose	Number of Trips	Trip Purpose	Number of Trips
Work	7,198	**Social/Recreational**	744
Work Related	216	**Personal Business**	883
School	2,943	**Returning Home**	56
Pick up/ Drop off passengers	476	**Other**	24
Shopping	2,252	**Grand Total**	14,792

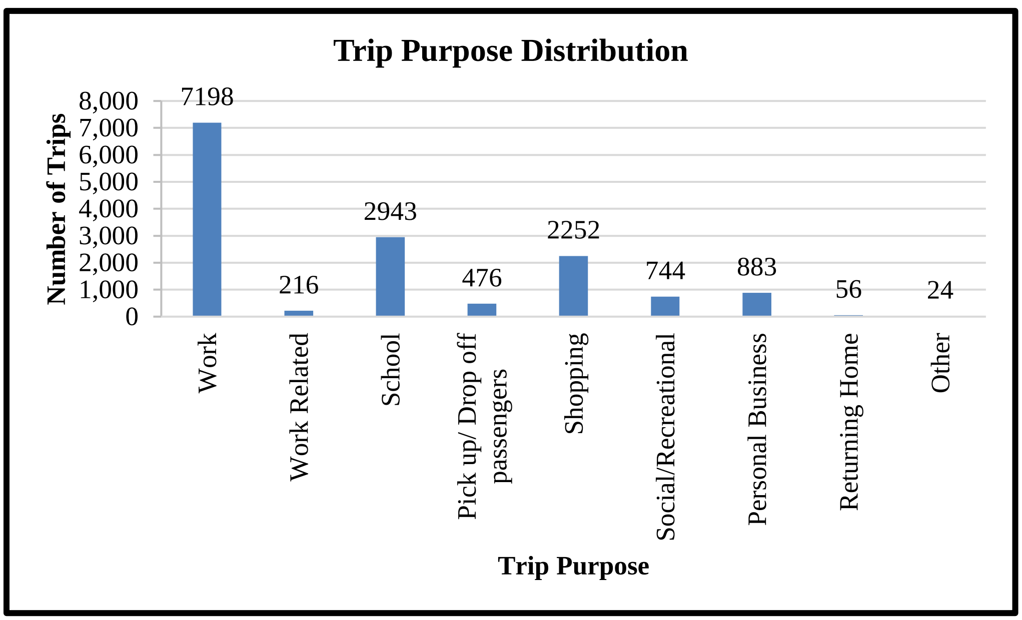

Figure A-6 Trip purpose distribution in dataset

Table A- 7 Method of travel distribution in dataset

Method of Travel	Number of Trips
Auto-Drive	9,341
Auto-Passenger	1,908
London Transit	1,268
Chartered Bus	26
School Bus	904
Taxi	52
Motorcycle	2
Bicycle	69
Walked all the way	1,205
Other	17
Grand Total	14,792

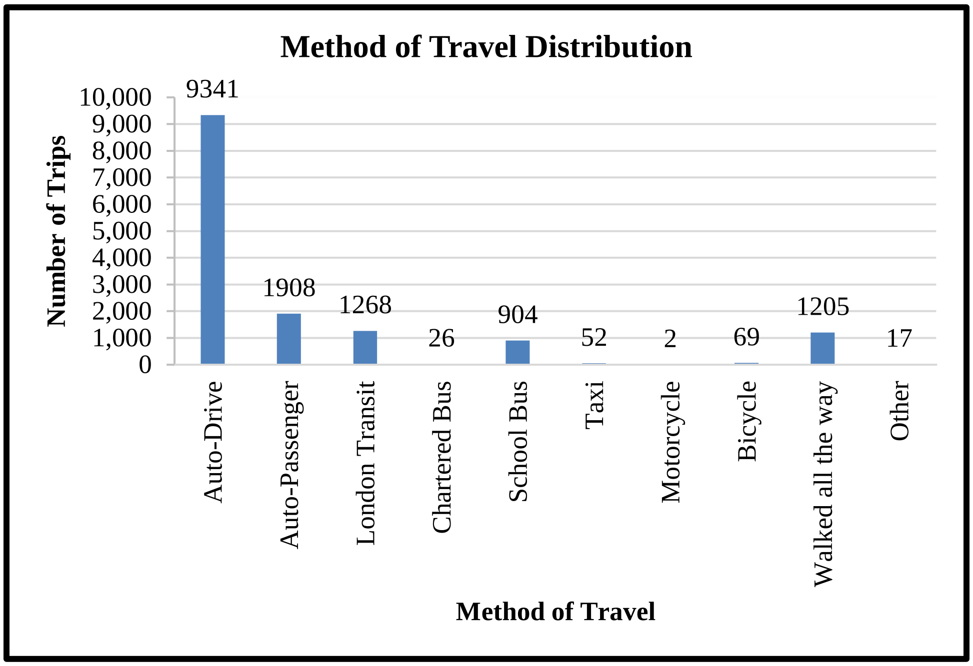

Figure A- 7 Method of travel distribution in dataset

Appendix B: Syntax for NLOGIT 5.0

Work Trip Generation Models

1) Linear Regression Model:
```
REGRESS; (1)
LHS = WTrip; (2)
RHS = One, Age2, Age3, Age4, Age5, Males, Vehicles, Dummy1, Dummy2, Dummy3,
      Dummy4$ (3)
```

2) Ordered Logit Model:
```
ORDERED; (4)
LHS = WTrip; (2)
RHS = One, Age2, Age3, Age4, Age5, Males, Vehicles, Dummy1, Dummy2, Dummy3,
      Dummy4$ (3)
```

3) Poisson Model:
```
POISSON; (5)
LHS = WTrip; (2)
RHS = One, Age2, Age3, Age4, Age5, Males, Vehicles, Dummy1, Dummy2, Dummy3,
      Dummy4$ (3)
```

4) Negative Binomial Model:
```
NEGBIN; (6)
LHS = WTrip; (2)
RHS = One, Age2, Age3, Age4, Age5, Males, Vehicles, Dummy1, Dummy2, Dummy3,
      Dummy4$ (3)
```

(1) Command to model a linear regression model

(2) Dependent variable, in this case it is the work trip generation count per household

(3) Independent variables considered in the model

(4) Command to model an ordered logit model

(5) Command to model a Poisson regression model

(6) Command to model a negative binomial model

Non-Work Trip Generation Models

1) Linear Regression Model:

```
REGRESS; (1)
LHS = NWTrip; (2)
RHS = One, Age2, Age3, Age4, Age5, Dummy5, Dummy6, Dummy7, Dummy8,
      SocialDummy, SocialxFemale$ (3)
```

2) Ordered Logit Model:

```
ORDERED; (4)
LHS = WTrip; (2)
RHS = One, Age2, Age3, Age4, Age5, Dummy5, Dummy6, Dummy7, Dummy8,
      SocialDummy, SocialxFemale$ (3)
```

3) Poisson Model:

```
POISSON; (5)
LHS = WTrip; (2)
RHS = One, Age2, Age3, Age4, Age5, Dummy5, Dummy6, Dummy7, Dummy8,
      SocialDummy, SocialxFemale$ (3)
```

4) Negative Binomial Model:

```
NEGBIN; (5)
LHS = WTrip; (2)
RHS = One, Age2, Age3, Age4, Age5, Dummy5, Dummy6, Dummy7, Dummy8,
      SocialDummy, SocialxFemale$ (3)
```

(1) Command to model a linear regression model

(2) Dependent variable, in this case it is the non-work trip generation count per household

(3) Independent variables considered in the model specification

(4) Command to model an ordered logit model

(5) Command to model a Poisson regression model

(6) Command to model a negative binomial model

Destination Type Model

Multinomial Logit Model:

NLOGIT; [1] CHOICES = Food, Medicine, Other; [2]

LHS = CHOICE; [3]

MODEL: [4]

U(Food) = Const1 + GroceryStore*Grocery + Walmart*Walmart + ln(Dist. to HWY)1*LnHwy + Age3xFemale*Age3Fem + VDLxFemale*VDLxFem/

U(Medicine) = Const2 + Shoppers Drug Mart*Shoppers Drug Mart + ln(Dist. to HWY)2*LnHwy + Age2*Age2/

U(Other) = DepartmentStore*Department + ln(Dist. to HWY)3*LnHwy $

Trip Distribution Final Micro-Models (MNL)

1) 5% Sample - Unconstrained Model

NLOGIT; [1]

LHS = CHOICE, COUNT; [5]

RHS = GroceryStore, DepartmentStore, Walmart, Costco, ShoppingMall, Shoppers Drug Mart, Other Pharmacies, ln(DistToHwy), ln(DistToCBD), MTW, Age(<20)xShoppers Drug Mart, Age(20-49) xDepartmentStore, Age(50-64)xCostco, Age(65+)xWalmart, FemalexShopping Mall, EmployedxDepartment Store, HouseholdSizexWalmart, HouseholdSizexCostco$ [6]

2) 5% Sample - Constrained Model

NLOGIT; [1]

LHS = CHOICE, COUNT; [5]

RHS = GroceryStore, DepartmentStore, Walmart, Costco, ShoppingMall, Shoppers Drug Mart, OtherPharmacies, ln(DistToHwy), ln(DistToCBD), MTW xGroceryStore, MTW xDepartmentStore, MTW xShoppingMall, MTW xShoppers Drug Mart, MTW xOtherPharmacies, Age(<20)xShoppers Drug Mart, Age(20-34)xOtherPharmacies, Age(20-49)xWalmart, Age(35-49)xGroceryStore, Age (35-49)xShoppers Drug Mart, Age(50-64)xCostco, Age(65+)x DepartmentStore, Employedx GroceryStore, HouseholdSizexGroceryStore, HouseholdSizexDepartmentStore, HouseholdSizexWalmart$ [6]

[1] Command to model a MNL model

[2] Definition of the alternatives considered

[3] Dependent variable, in this case it is the destination type choice

[4] Utility function with β parameter estimates

[5] Dependent variable, in this case it is the micro-destination choice sets

[6] Independent variables considered in the model specification

Trip Distribution Final Macro-Models (MNL)

1) 5% Sample - Unconstrained Model
NLOGIT; [1]
LHS = CHOICE, COUNT; [2]
RHS = GroceryStore, Walmart, Shoppers Drug Mart, Other Pharmacies, DepartmentStore, VarietyStore, HardwareStore, AutoandHomeSupplyStore, WomenClothingStore, ShoeStore, FurnitureStore, ShoppingMall, AlcoholicBevarages, CommercialBanks, ln(DistToCBD), Femalexln(DistanceToCBD), Age(20-49)xShoppers Drug Mart, Age(20-34)xDepartmentStore, Age(20-34)xAutoandHomeSupplyStore, Age(35-49)xWalmart, Age(35-49)xHardware Store, Age(35-49)xWomenClothingStore, Age(35-49)xShoppingMall, Age(35-49)x AlcoholicBeverages, Age(65+)xOtherPharmacies, Age(65+)xDepartmentStore$ [3]

2) 5% Sample - Constrained Model
NLOGIT; [1]
LHS = CHOICE, COUNT; [2]
RHS = GroceryStore, Walmart, Shoppers Drug Mart, Other Pharmacies, DepartmentStore, VarietyStore, HardwareStore, AutoandHomeSupplyStore, WomenClothingStore, ShoeStore, FurnitureStore, ShoppingMall, AlcoholicBevarages, CommercialBanks, ln(DistToCBD), Femalexln(DistanceToCBD), MTW, Femalex MTW, Age(20-49)xShoppers Drug Mart, Age(20-34)x DepartmentStore, Age(20-34)xAutoandHomeSupplyStore, Age(35-49)xWalmart, Age(35-49)xHardwareStore, Age(35-49)xWomenClothingStore, Age(35-49)xShopping Mall, Age(35-49)xAlcoholicBeverages, Age(65+)xOtherPharmacies, Age(65+)x DepartmentStore$ [3]

[1] Command to model a MNL model

[2] Dependent variable, in this case it is the macro-destination choice sets (TAZ)

[3] Independent variables considered in the model specification

VITA AUCTORIS

NAME:	Rahaf Husein
PLACE OF BIRTH:	Amman, Jordan
YEAR OF BIRTH:	1994
EDUCATION:	Vincent Massey Secondary School, Windsor, ON, 2011
	University of Windsor, B.Sc., Windsor, ON, 2015
	University of Windsor, M.Sc., Windsor, ON, 2017